Journal of Cheminformatics 2012, Volume 4 Suppl 1
http://www.jcheminf.com/supplements/4/S1

Journal of Cheminformatics

MEETING ABSTRACTS Open Access

7th German Conference on Chemoinformatics: 25 CIC-Workshop

Goslar, Germany. 6-8 November 2011

Edited by Frank Oellien, Uli Fechner and Thomas Engel

Published: 1 May 2012

These abstracts are available online at http://www.jcheminf.com/supplements/4/S1

INTRODUCTION

A1

GCC2011 – 25 years of computational chemistry meetings

Frank Oellien[1*], Uli Fechner[2], Thomas Engel[3]
[1]GDCh-CIC Division Chair, Intervet Innovation GmbH – MSD Animal Health, Zur Propstei, 55270 Schwabenheim, Germany; [2]GDCh-CIC Division Board Member, Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Trakehner Str. 7-9, 60487 Frankfurt, Germany; [3]GDCh-CIC Division Co-Chair and Conference Chair, Fakultät für Chemie und Pharmazie, Universität München, Butenandtstr. 5-13, 81377 München, Germany
E-mail: Frank.Oellien@msd.de
Journal of Cheminformatics 2012, **4(Suppl 1):**A1

25 years ago members of the Chemistry-Information-Computers (CIC) division of the German Chemical Society (GDCh) [1] realized that the usage of computers will play a major role in the processing of chemical information and that computational methods will have a large impact on chemical research approaches. At that time, Computational Chemistry was not yet established as a research field. Thus, scientists working in this area were typically chemists from all realms of chemistry who happened to have an interest in computers. To address the initial issues in the field of Computational Chemistry in a collaborative manner, the workshop *"Software-Entwicklung in der Chemie"* (software development in chemistry) — later renamed the CIC-Workshop — was established. The foundations were laid in this annual workshop and many projects and scientific outcomes originated from it in the years thereafter. While the initial workshops focused on the implementation of chemical databases, other topics, such as structure elucidation, structure representation and data mining, gained importance over the following years. The scientific network became more and more international over the intervening years, so much so that the CIC board decided in 2005 to change the, up to that point, German workshop into an international conference. The **7th German Conference on Chemoinformatics** (GCC2011) was held from the 6th to the 8th of November 2011 in Goslar, Germany. The CIC division invited the chemoinformatics and molecular modeling community to the GCC2011 to celebrate the 25th anniversary of the CIC-workshop.
The conference focused on recent developments and trends in the fields of
Chemoinformatics and Drug Discovery
Chemical Information, Patents and Databases
Molecular Modeling
Computational Materials Science and Nanotechnology
As always, contributions from other research areas of Computational Chemistry were also welcome.
Despite the recent major changes in the pharmaceutical industry and the resulting decrease in research, the number of participants was comparable to the German Conference on Chemoinformatics in 2010. The international character of the conference was even more pronounced than in the preceding years due to the 149 participants from 18 countries (Figure 1, 2). Following the tradition, the conference was opened up by the *"Free-Software-Session"* and the *"Chemoinformatics Market Place"* on Sunday afternoon. Four Open Source projects — Travis, Knime, ParadocS and DebiChem — were presented in the *"Free-Software-Session"* and three preconference workshops were given by the companies Chemical Computing Group, Tripos and Xemistry. The first day of the conference was concluded by dinner and an evening lecture by Johann Gasteiger (*"25 Years of CIC – Achievements and Future Goals"*) both of which took place in the ore mine of Rammelsberg.
The program of the following two days included plenary lectures from six well-known keynote speakers from industry and academia (Oliver Kohlbacher, University of Tübingen, Germany; Colin Groom, CCDC, Cambridge, UK; Eva Rauls, University of Paderborn, Germany; Colin Batchelor, RCS, Cambridge, UK; Herbert Köppen, Boehringer Ingelheim, Germany and Xavier Barril, University of Barcelona, Spain), as well as 17 lectures and 71 poster presentations.
A special highlight of the conference was the FIZ-CHEMIE-Berlin awards (Figure 3). The CIC division awards this prize every year to the best diploma thesis and the best PhD thesis in the field of Computational Chemistry. The prize for the PhD thesis was awarded to Dr. Volker Hähnke from the group of Prof. Gisbert Schneider, ETH Zurich for his dissertation *"Text-based Similarity Searching for Hit- and Lead-Candidate Identification"*. The award for the best diploma thesis was given to Daniel Moser from the group of Jun. Prof. Eugen Proschak, University of

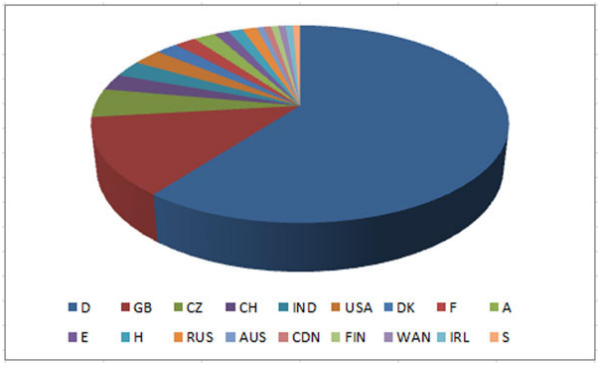

Figure 1(abstract A1) 149 participants from 18 countries and 5 continents attended the GCC2011.

Figure 2(abstract A1) Participants of the 7th German Conference on Chemoinformatics (GCC2011), November 6 – 8, 2011 in Goslar, Germany.

Figure 3(abstract A1) FIZ CHEMIE Berlin Awards 2011: from left to right, Rene de Planque (Head of the FIZ CHEMIE Berlin), Volker Hähnke (FIZ CHEMIE Berlin awardee, dissertation prize; NCBI, Bethesda, USA), Daniel Moser (FIZ CHEMIE Berlin awardee, master thesis prize; University of Frankfurt, Germany) and Frank Oellien (Chair of the GDCh CIC division).

Frankfurt for his excellent master thesis "*Design of Dual Ligands Using Excessive Pharmacophore Query Alignment*". Due to the closedown of the FIZ CHEMIE Berlin, these prizes were awarded for the last time in 2011. To continue its support of young German scientists in the future, starting with the GCC2012, the CIC division will endow the ***"CIC Advancement Award for Computational Chemistry"***.

The conference ended on Tuesday evening with a conference dinner and a speaker who was announced as a "special guest from MIT". The special guest turned out to be Thomas Fraps, a magician who presented a very entertaining show. In the end he left a pleased and puzzled audience who could not find the answers to his "scientific experiments". A lot of problems in Computational Chemistry have been solved in the last 25 years, yet the road ahead is full of challenging issues waiting to be tackled.

References

1. **GDCh CIC Division.** [http://www.gdch.de/cic].
2. **FIZ CHEMIE Berlin Award.** [https://www.gdch.de/netzwerk-strukturen/fachstrukturen/chemie-information-computer-cic/cic-foerderpreis-fuer-computational-chemistry.html].

ORAL PRESENTATIONS: FREE SOFTWARE SESSION

F1

TRAVIS - a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories

Martin Brehm*, Barbara Kirchner
Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Universität Leipzig, Leipzig, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**F1

We present TRAVIS ("TRajectory Analyzer and VISualizer"), a free program package for analyzing and visualizing Monte Carlo and molecular dynamics trajectories [1]. The aim of TRAVIS is to collect as many analyses as possible in one program, creating a powerful tool and making it unnecessary to use many different programs for evaluating simulations. This should greatly rationalize and simplify the workflow of analyzing trajectories. TRAVIS is written in C++, open-source freeware and licensed under the terms of the GNU General Public License (GPLv3). It is easy to install (platform independent, no external libraries) and easy to use. On this poster, we present some of the algorithms that are implemented in TRAVIS - many of them widely known for a long time, but some of them also to appear in literature for the first time. All shown analyses only require a standard MD trajectory as input data.

Reference

1. Brehm M, Kirchner B: **TRAVIS - A free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories.** *J Chem Inf Model* 2011, **51**:2007-2023, doi:10.1021/ci200217w.

F2

What's new in KNIME?

Thorsten Meinl
University of Konstanz, Box 712, 78457 Konstanz, Germany
E-mail: Thorsten.Meinl@uni-konstanz.de
Journal of Cheminformatics 2012, **4(Suppl 1):**F2

KNIME (Konstanz Information Miner, [1]) is a user-friendly and comprehensive open-source data integration, processing, analysis, and exploration platform. From day one, KNIME has been developed using rigorous software engineering practices and is used by professionals in both industry and academia in over 60 countries.
In the presentation we will show some of the new features in KNIME 2.4 and give an outlook to KNIME 2.5. We also present the KNIME Community Contributions [2], where research groups can easily provide their KNIME extensions to the community.

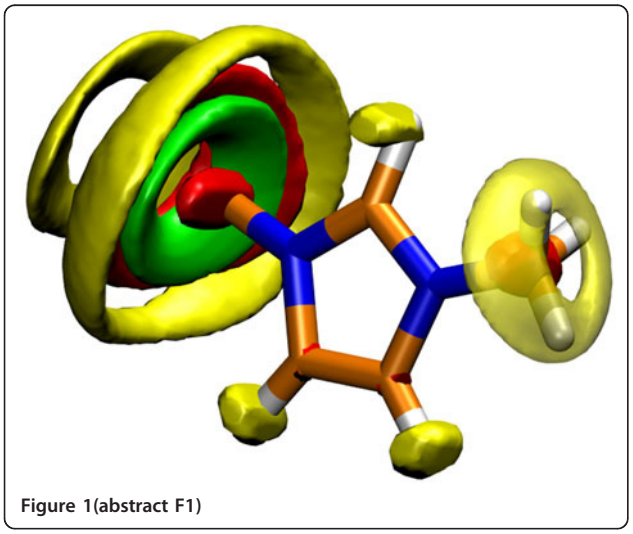

Figure 1(abstract F1)

References

1. [http://www.knime.org/].
2. [http://tech.knime.org/community].

F3

ParaDockS – an open source framework for molecular docking

Martin Pippel[1*], Michael Scharfe[1*], Renè Meier[2], Wolfgang Sippl[1]
[1]Institute of Pharmacy, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), 06120, Germany; [2]Institute of Biochemistry, University of Leipzig, Leipzig, 04103, Germany
E-mail: martin.pippel@pharmazie.uni-halle.de
Journal of Cheminformatics 2012, **4(Suppl 1):**F3

The generation and evaluation of molecular complexes in order to predict binding modes of protein ligand complexes, is still a challenging task. Many different algorithms and approaches try to solve the molecular docking problem. Within this work we introduce the open source docking platform ParaDockS (Parallel Docking Suite) [1], that allows the convenient incorporation of existing and new approaches either to describe ligand-receptor interaction or to search for native poses.
The goal of this work pursues the line of thought, that ParaDockS is open source. We want to initiate a highly dynamic community, so that both users and developers contribute with their knowledge and experience to take this very interesting field one step further.
The framework, written in C++, is highly modular designed. This offers an easy access to developers to work only on their area of expertise. ParaDockS provides a functionality to analyze, process and store different kind of input structures.
Another focus is set on the implementation of scoring functions. Beside implemented scoring functions PScore, PMF04 and Drugscore, ParaDockS offers the possibility to derive and apply target specific scoring functions. Therefore, we implemented a workflow to derive target class specific PMF atom-pair potentials. This involves the preparation of complex structures to get an adequate structural database. The definition of atom types and the derivation of pair potentials can completely be accomplished in the upcoming version of ParaDockS. The resulting atom pair potentials can be directly used as scoring function for docking or rescoring approaches. The whole workflow was applied and validated for kinases, but is applicable to every target class with enough structural data.
One of the latest improvements involves an interface to the pharmacophor modeling program LigandScout [3]. This offers the big advantage to setup the docking with a great graphical interface, as well the created pharmacophores can be used to guide the docking process.

References

1. Meier, Pippel, Brandt, Sippl, Baldauf: **ParaDockS – A framework for Molecular Docking with Population-Based Metaheuristics.** *J Chem Inf Model* 2010, **50**:879-889.
2. Muegge I: **PMF Scoring Revisited.** *J Med Chem* 2006, **49**:5895-5902.
3. Wolber G, Langer T: **LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters.** *J Chem Inf Model* 2005, **45**:160-169.

F4

Packaging free software chemistry programs in Debian GNU/Linux: past, present and future

Michael Banck[1,2]
[1]Department of Chemistry, Technische Universität München, Lichtenbergstrasse 4, D-85747 Garching, Germany; [2]Debian Project
E-mail: michael.banck@ch.tum.de
Journal of Cheminformatics 2012, **4(Suppl 1):**F4

Started in 1993, the Debian project is one of the oldest Free Software projects. Due to its volunteer nature, specialists from all fields contribute to the Debian GNU/Linux distribution, which includes more than 30000 packages. The Debian packaging policy, its advanced package management system and the conservative release process lead to a stable basis which is ideal for customized environments like scientific research.
The Debichem project [1] has been packaging and maintaining chemical software compliant with the Debian Free Software Guidelines (DFSG) [2] since 2006. Currently, 35 Free Software package are directly maintained

by the Debichem team and 10 more are part of Debichem but maintained by others.
At the core of Debichem are the cheminformatics packages OpenBabel [3], CDK [4] and RDKit [5]. They provide file format conversion, 3D coordinate generation, molecular descriptors and fingerprints, stereochemistry prediction, conformation generation and searching, forcefields and more.
Besides those, a variety of 2D/3D visualization and molecular modelling programs, as well as ab initio, semi-empirical and molecular dynamics codes are packaged by the Debichem in Debian.
Future work will include packaging of cinfony [6], a python module which presents a common API over OpenBabel, CDK and RDKit.

References

1. [http://alioth.debian.org/projects/debichem].
2. [http://www.debian.org/social_contract#guidelines].
3. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR: **Open Babel: An Open Chemical Toolbox.** *J Cheminf* 2011, **3**:33.
4. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen EL: **The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics.** *J Chem Inf Comput Sci* 2003, **43(2)**:493-500.
5. [http://rdkit.org/].
6. O'Boyle NM, Hutchison GR: **Cinfony – combining Open Source cheminformatics toolkits behind a common interface.** *Chem Cent J* 2008, **2**:24.

ORAL PRESENTATIONS

O1

25 years of CIC – achievements and future goals

Johann Gasteiger[1,2]
[1]Computer-Chemie-Centrum, University of Erlangen Nuremberg, Germany; [2]Molecular Networks GmbH, Henkestr. 91, D-91052 Erlangen, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**O1

25 years ago several persons within the German Chemical Society realized that computer methods will increasingly play a major role in the processing of chemical information. This led to the foundation of the Division "Chemie-Information-Computer (CIC)" and the establishment of a Workshop "Software-Entwicklung in der Chemie", Parallel to this meeting the "Molecular Modeling Workshop" was initiated covering more the fields of molecular modeling and drug design. Over the years the CIC Workshop embraced the English language and metamorphosed into a European endeavor.
From the very beginning scientists of all disciplines of chemistry participated in this Workshop. Initially, a major emphasis was put on the building of databases paralleling the work on the Beilstein, the Gmelin, the SpecInfo and the ChemInform databases. However, already in the early phases of this Workshop research was presented on converting chemical information into knowledge and using this knowledge in systems for automatic structure elucidation or the design of organic syntheses.
The Workshop certainly was instrumental in introducing Chemoinformatics in Germany. Many problems in Chemoinformatics were tackled and solved. However, it must be realized that by far not all problems have been solved. Furthermore, the recognition of Chemoinformatics as a scientific discipline still leaves something to be desired. These topics will be discussed and approaches for their improvement will be outlined.

O2

CADDSuite – a workflow-enabled suite of open-source tools for drug discovery

Oliver Kohlbacher
Applied Bioinformatics, Center for Bioinformatics, University of Tübingen, Germany
E-mail: oliver.kohlbacher@uni-tuebingen.de
Journal of Cheminformatics 2012, **4(Suppl 1):**O2

We present an open-source suite of tools for computer-aided drug design, CADDSuite, built upon our molecular modeling framework BALL. CADDSuite provides a wide range of tools for structure preparation, docking, QSAR and related topics.
IMGDock is a novel docking tool combining heuristic search strategies and a grid-based scoring function. We demonstrate that it yields results comparable other current docking tools on popular docking benchmark sets. It can be easily combined with TagRes, a tool for target-specific rescoring that allows a more accurate estimation of binding free energies within a specific target family.
All tools of the CADDSuite, including IMGDock and TagRes, are open-source software (under the GNU Public License). Tools of the CADDSuite share a common binary interface, a common data exchange format and thus easily integrate into distributed computing environments. We show how they can be used from KNIME and how KNIME workflows can then be executed on distributed resources, in particular Galaxy and WS-PGRADE for grid computing.

O3

A computational method to facilitate structure-activity relationship transfer

Anne Mai Wassermann*, Jürgen Bajorath
Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, 53113, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**O3

The transfer of SAR information from one chemical series to another is a challenging task in medicinal chemistry. However, this process is highly relevant for lead optimization because it is not uncommon that a compound series displaying a promising SAR has other liabilities, e.g. toxic side effects. In such cases, one would ideally like to build upon prior knowledge and utilize the available SAR information for the optimization of an alternative chemotype. For this purpose, alternative molecular core structures must be identified where corresponding chemical substitutions yield comparable SAR trends (consistent with a conserved mechanism of action). To support this process, we have developed a data mining method that, for the first time, enables the identification of alternative analog series with different core structures but corresponding substitution patterns and comparable potency progression [1]. A scoring scheme is utilized to evaluate the possibility of a successful SAR transfer to another series. In addition, a graphical representation has been designed to simultaneously monitor potency changes for multiple analog series as a consequence of defined substitutions. The method is also applicable to explore other aspects of SAR transfer such as comparative learning from multiple compound series with varying degrees of chemical exploration, which often results in suggestions for the design of new analogs. Furthermore, the approach has been applied to systematically assess SAR transfer potential in public compound data [1].

Reference

1. Wassermann AM, Bajorath J: **A Data Mining Method to Facilitate SAR Transfer.** *J Chem Inf Model* 2011, **51(8)**:1857-1866.

O4

Aligning chemical structure diagrams with local search

Matthias Hilbig*, Matthias Rarey
Zentrum für Bioinformatik, Universität Hamburg, 20146, Germany
E-mail: hilbig@zbh.uni-hamburg.de
Journal of Cheminformatics 2012, **4(Suppl 1):**O4

Chemists working in biomolecular application projects are usually looking at many related molecules (e.g. results of a virtual screening run, lead series development or library design). For a convenient visual analysis of this data it is essential that differences between molecules are easily detectable. This can be quite difficult if structural similarities are not taken into account while creating molecule structure diagrams. We present a method for generating globally aligned structure diagrams for two molecules following IUPAC standards [1]. Using a set of three coordinate transform operations (ring system flipping, chain flipping and substituent exchange) all correct and overlap-free layouts can be enumerated. If the number of possible layouts is too large, a heuristic is used to iterate through a smaller subspace. Subsequently all candidate layouts are scored with several different terms describing the quality of the layout (number of collisions, stretching of chains...) as well as the relationship between molecules (similarity to a template) and the one with the highest score is chosen. Scoring functions and similarity measures are easily interchangeable and the whole process is

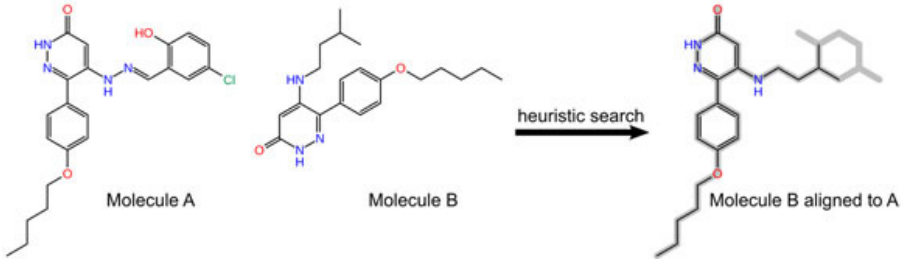

Figure 1(abstract O4)

fast enough for interactive use. The whole alignment process is verified by calculating the RMSD between aligned and nearest template coordinates. For validation, the new method is applied to many clusters of related molecules from the PubChem compound library. In summary, we have developed a novel SDG algorithm which is of great help for the daily tasks of a modeller by drawing small, related molecules consistently.

References

1. Brecher J: **Graphical representation standards for chemical structure diagrams (IUPAC Recommendations 2008).** *Pure Appl Chem* 2008, **80**:277-410.
2. Fricker PC, Gastreich M, Rarey M: **Automated Drawing of Structural Molecular Formulas under Constraints.** *J Chem Inf Comput Sci* 2004, **44**:1065-1078.
3. Boissonnat J, Cazals F, Flötotto J: **2D - structure drawings of similar molecules.** *Graph Drawing* 2001, **1984**:89-93.
4. Clark AM: **2D depiction of fragment hierarchies.** *J Chem Inf Model* 2010, **50**:37-46.

O5

Chemoinformatics in drug development

Colin R Groom
Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK
E-mail: groom@ccdc.cam.ac.uk
Journal of Cheminformatics 2012, **4(Suppl 1):**O5

It would be unimaginable to prosecute a drug discovery program without applying appropriate chemoinformatics analyses. In recent years a focus on target affinity and activity has been complemented by techniques to address physico-chemical properties such as lipophilicity and solubility, biological properties such as absorption, distribution, metabolism, elimination and toxicity. As such, a rounded package of studies can be performed to help in the generation and selection of molecules as clinical development candidates. Medicinal chemists are often well-served by their computational chemistry colleagues. Not so the development chemist.
Having successfully produced a clinical candidate the attention of chemoinformaticians in the pharmaceutical industry usually turns to the next molecule and scant regard is given to the contributions that can be made as a candidate molecule progresses towards becoming part of a drug substance.
This presentation will highlight the opportunities for the application of chemoinformatics techniques to the area of pharmaceutical materials science, a critical and complex phase in the creation of a drug.

O6

Winnow based identification of potent hERG inhibitors *in silico*: comparative assessment on different datasets

Richard L Marchese Robinson[1*], Robert C Glen[1], John BO Mitchell[2]
[1]Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK;
[2]EaStCHEM School of Chemistry and Biomedical Sciences Research Centre, University of St Andrews, North Haugh, St Andrews, KY16 9ST,UK
E-mail: rlm46@cam.ac.uk
Journal of Cheminformatics 2012, **4(Suppl 1):**O6

Due to the potentially lethal effects of potent hERG inhibition, *in silico* approaches which can identify potent (IC_{50} < 1 μM) inhibitors are of considerable interest to the pharmaceutical industry [1]. We present recent work [2] in which *in silico* binary classifiers were trained to discriminate potent inhibitors from compounds exhibiting weaker (IC_{50} ≥ 1 μM) inhibition. Initial models were based on a version of the memory efficient Winnow algorithm [3]. These initial models were generated using various descriptor sets. The descriptor set yielding the best cross-validated initial Winnow model was used to build models using each of Winnow, Random Forest and Support Vector Machine. Analysis of the contributions of different substructural and physiochemical features in the final Winnow models indicates they may be interpreted, albeit with caution. All final models were externally validated, with no algorithm consistently outperforming the others. These approaches were directly compared, on various datasets, to those proposed by Thai and Ecker [4] and by Dubus et al. [5]. The results indicate that the Winnow models are competitive with earlier approaches proposed in the literature. The findings also emphasise a potential difficulty when seeking to estimate the predictive power of *in silico* models on small quantities of data: model performance may vary considerably, particularly when training and validating on different datasets.

References

1. Gavaghan CL, Hasselgren Arnby C, Blomberg N, Strandlund G, Boyer S: **Development, interpretation and temporal evaluation of a global QSAR of hERG electrophysiology screening data.** *J Comput-Aided Mol Des* 2007, **21**:189-206.
2. Marchese Robinson RL, Glen RC, Mitchell JBO: **Development and Comparison of hERG Blocker Classifiers: Assessment on Different Datasets Yields Markedly Different Results.** *Mol Inf* 2011, **30**:443-458.
3. Nigsch F, Mitchell JBO: **How To Winnow Actives from Inactives: Introducing Molecular Orthogonal Sparse Bigrams (MOSBs) and Multiclass Winnow.** *J Chem Inf Model* 2008, **48**:306-318.
4. Thai KM, Ecker GF: **A binary QSAR model for classification of hERG potassium channel blockers.** *Bioorg Med Chem* 2008, **16**:4107-4119.
5. Dubus E, Ijjaali I, Petitet F, Michel A: **In Silico Classification of hERG Channel Blockers: a Knowledge-Based Strategy.** *ChemMedChem* 2006, **1(6)**:622-630.

O7

Putting the available chemical space to the fingertips of our scientists

Jörg Degen[1*], Mark Rogers-Evans[2], Daniel Stoffler[1]
[1]Cheminformatics & Statistics, F. Hoffmann-La Roche Ltd., Basel, Switzerland;
[2]Discovery Chemistry, F. Hoffmann-La Roche Ltd., Basel, Switzerland
E-mail: joerg.degen@roche.com
Journal of Cheminformatics 2012, **4(Suppl 1):**O7

One of the most important and frequently occurring questions to answer in the course of a drug design project is which compounds to synthesize next. If it were possible to search virtually all available molecules via one interface, new chemical design suggestions and ideas could be followed up more quickly, thus drastically reducing the cycle time needed for hypotheses testing. Furthermore, if there was a platform available that offered sufficient flexibility for combining various constraints and for handling fuzziness in chemical substructures, promising and non-obvious

chemical variations could be identified more quickly and easily, resulting in a more efficient exploration of chemical space.

As a result of a close and long standing collaboration between Discovery Chemistry and Cheminformatics, we have established a global system that allows every researcher at Roche to access all available small molecules that can either be obtained from internal sources or via external vendors. For the more than 10 Million physically accessible compounds, the associated structural and physicochemical properties are included as well as information on existing quantities, locations, prices and availability. Our tailored designed web-based interface and underlying query engine allows for the identification of compounds using various constraints and simultaneously provides utmost flexibility for combining different search criteria. Similarity and particularly fuzzy search capabilities facilitate an intuitive navigation of the available chemical space. The whole system is very fast so that query results are returned almost instantly, allowing for a quick and interactive drill-down to manageable sets of compounds.

The system is being used globally by a diverse user community and applied for addressing many different kinds of questions, for example, for fast identification of hit expansion candidates, exploration of building block libraries, probing of our in-house screening library, compilation of screening compound subsets, or finding appropriate scaffold replacements. To our knowledge, there is no other platform available that provides similar flexibility, performance and usability for accessing diverse chemicals at such speed.

During the presentation we will outline the basic concepts and architecture of the system and point out some key challenges that we were facing during its development. In addition, we will give some examples on how we integrated and search for recent Roche-ETH driven developments of potent, novel, alternative building blocks in medicinal chemistry [1].

Reference

1. Burkhard JA, Wuitschik G, Rogers-Evans M, Müller K, Carreira EM: **Oxetanes as Versatile Elements in Drug Discovery and Synthesis.** *Angew Chem Int Ed Engl* 2010, **49**:9052-9067.

O8

Quantifying intrinsic chemical reactivity of molecular structural features for protein binding and reactive toxicity, using the MOSES chemoinformatics system

Johannes AH Schwöbel[1*], Bruno Bienfait[1], Johann Gasteiger[1], Thomas Kleinöder[1], Jörg Marusczyk[1], Oliver Sacher[1], Christof H Schwab[1], Aleksey Tarkhov[1], Lothar Terfloth[1], Mark TD Cronin[2]
[1]Molecular Networks GmbH, Henkestraße 91, D-91052 Erlangen, Germany; [2]School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK
E-mail: schwoebel@molecular-networks.com
Journal of Cheminformatics 2012, **4(Suppl 1):**O8

Covalent binding of xenobiotic compounds to endogenous biomolecular sites, e.g. protein residues, leads to potentially irreversible toxic effects such as enhanced acute toxicity or skin sensitization [1]. This mechanistic knowledge provides the basis for the *in silico* prediction of these toxicities, as required by the EU REACH legislation and the EU Cosmetics Directive. A general toxicity prediction can be based on three consecutive steps [2]: (1.) *Identification of a potential reactive protein binding mechanism via a set of molecular structural features*. Those structural features can be encoded by the Chemical Subgraph Representation Markup Language (CSRML), that supports a flexible annotation of meta information, including physicochemical properties as annotations. (2.) *Confirmation and quantification of* (*bio*)*chemical reactivity*. The potential for a chemical to be reactive can be captured by mechanistically based QSAR models. This intrinsic reactivity is calculated rapidly by descriptors of the chemoinformatics platform Molecular Structure Encoding System (MOSES) [3]. It represents electronic, steric and resonance effects in chemical structures. The performances obtained by these reactivity models are close to time-consuming quantum chemical reactivity calculations, e.g., *se* = 0.44 *versus* 0.40 for glutathione adduct formation via Michael addition, comparing predicted values to an experimental reactivity dataset [1]. (3.) *Establishing a relationship between calculated reactivity and toxicity*. The predicted intrinsic reactivity values were linked to the computational prediction for different modes of toxic action, with good correlations between predicted and *in vitro* toxicity (up to r^2=0.86).

The combined use of structural information and computed reactivity could assist in the non-animal based risk assessment of chemicals for regulatory purposes and in the application of integrated testing strategies (ITS).

The research has received funding from the EU FP7 COSMOS Project (grant agreement n° 266835) and financing from COLIPA.

References

1. Schwöbel JAH, *et al*: **Measurement and estimation of electrophilic reactivity for predictive toxicology.** *Chem Rev* 2011, **111**:2562-2596.
2. Schwöbel JAH, *et al*: **Application of a computational model for Michael addition reactivity in the prediction of toxicity to *Tetrahymena pyriformis*.** *Chemosphere* 2011, **85(6)**:1066-1074.
3. **MOSES**, is a C++ chemoinformatics software library that is developed and maintained by Molecular Networks GmbH, Erlangen, Germany (http://www.molecular-networks.com/software/moses/).

O9

Non-continuum solvation using the EC-RISM method applied to predict tautomer ratios, pK_a and enantiomeric excess of alkylation reactions

Jochen Heil*, Roland Frach, Stefan M Kast
Technische Universität Dortmund, Dortmund, 44221, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**O9

The three-dimensional "reference interaction site model" (3D-RISM) integral equation theory is a statistical-mechanical approach to predict liquid state structural and thermodynamic features. It is based on approximate solute-solvent correlation functions to be computed on a 3D grid as a function of the interaction potential between the solute and the solvent sites, circumventing the need of costly sampling of explicit solvent degrees of freedom. In combination with quantum-chemical calculations within the embedded cluster (EC-)RISM framework [1] the theory allows for studying chemical reactions in solution with an accuracy not reached by traditional continuum solvation methods. In particular, it improves upon dielectric continuum solvation by taking solvent granularity into account and also provides a means towards physically cavity formation and dispersion free Energies without introducing artificial boundaries and empirically fitted radii.

We outline the general framework and show application examples from pK_a and tautomeric ratio estimation [2] as well as enantiomeric excess prediction for stereoselective alkylation reactions in organic solvent.

References

1. Kloss T, Heil J, Kast SM: **Quantum Chemistry in Solution by Combining 3D Integral Equation Theory with a Cluster Embedding Approach.** *J Phys Chem B* 2008, **112**:4337-4343.
2. Kast SM, Heil J, Güssregen S, Schmidt KF: **Prediction of tautomer ratios by embedded-cluster integral equation theory.** *J Comput-Aided Mol Des* 2010, **24**:343-353.

O10

Integrating logic-based machine learning and virtual screening to discover new drugs

Christopher R Reynolds*, Michael JE Sternberg
Imperial College, London, N19 3LA, UK
E-mail: cr308@imperial.ac.uk
Journal of Cheminformatics 2012, **4(Suppl 1):**O10

Investigational **N**ovel **D**rug **D**iscovery by **Ex**ample (INDDEx™) is a technology developed to guide hit to lead discovery by learning rules from existing active compounds that link activity to chemical substructure. INDDEx is based on Inductive Logic Programming [1], which learns easily interpretable *qualitative* logic rules from active ligands that give an insight into chemistry, relate molecular substructure to activity, and can be used to guide the next steps of drug design chemistry. Support Vector Machines weight the rules to produce a *quantitative* model of structure-activity relationships. Whereas earlier testing [2,3] was performed on single dataset examples, this talk presents the largest and fullest test of the method. The method was benchmarked on the Directory of Useful Decoys (DUD) datasets [4], using the same methodology described in the paper on the assessment of LASSO [5] and DOCK. For each of the DUD datasets, the known active ligands were mixed with all the decoy compounds in DUD, and the retrieval rates of INDDEx and DUD were measured when they

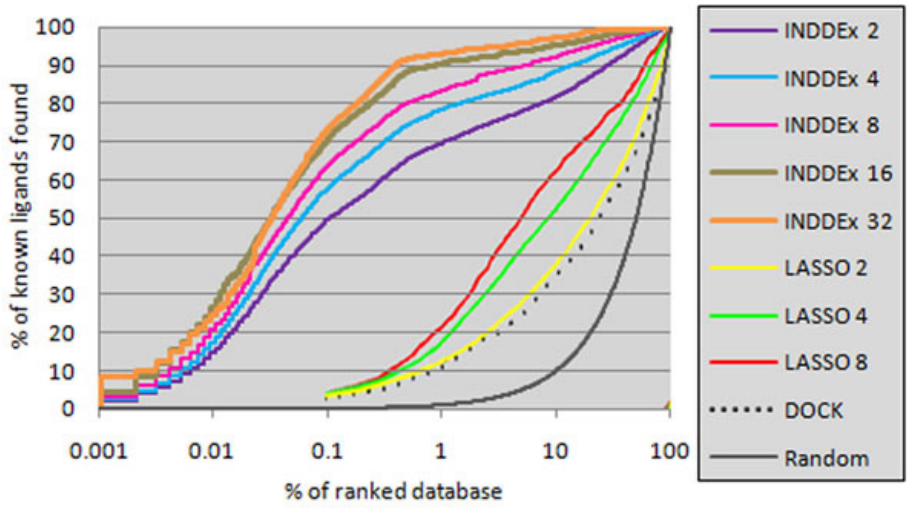

Figure 1(abstract O10) Recovery of actives in each of the DUD datasets from all decoys in the DUD, averaged across all 40 datasets.

were trained on 2, 4, and 8 of the known active ligands (Figure 2). Early retrieved compounds showed high topological differences to molecules used as training data, showing the strength of this method for scaffold hopping. This work was supported by a BBSRC case studentship with Equinox Pharma Ltd (http://www.equinoxpharma.com).

References

1. Muggleton SH: **Inductive logic programming.** *New Gen Comp* 1995, **13**:245-286.
2. Amini A, *et al*: **A Novel Logic-Based Approach for Quantitative Toxicology Prediction.** *J Chem Inf Model* 2007, **47**:998-1006.
3. Cannon EO, *et al*: **Support vector inductive logic programming outperforms the naive Bayes classifier and inductive logic programming for the classification of bioactive chemical compounds.** *J Comput Aided Mol Des* 2007, **21**:269-280.
4. Huang N, *et al*: **Benchmarking Sets for Molecular Docking.** *J Med Chem* 2006, **49**:6789-6801.
5. Reid D, *et al*: **LASSO-ligand activity by surface similarity order: a new tool for ligand based virtual screening.** *J Comput Aided Mol Des* 2008, **22**:479-487.

O11

Design of dual ligands using excessive pharmacophore query alignment

Daniel Moser*, Joanna Wisniewska, Steffen Hahn, Estel la Buscató, Franca-Maria Klingler, Janosch Achenbach, Bettina Hofmann, Dieter Steinhilber, Ewgenij Proschak
Institute of Pharmaceutical Chemistry, LiFF/OSF/ZAFES, Goethe University, Frankfurt am Main, Germany, Max-von-Laue Str.9, D-60438, Frankfurt/M, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**O11

Dual- or multi-target ligands have gained increased attention in the past years due to several advantages, including more simple pharmacokinetic and phamarcodynamic properties compared to a combined application of several drugs. Furthermore multi-target ligands often possess improved efficacy [1]. We present a new approach for the discovery of dual-target ligands using aligned pharmacophore models combined with a shape-based scoring. Starting with two sets of known active compounds for each target, a number of different pharmacophore models is generated and subjected to pairwise graph-based alignment using the Kabsch-Algorithm [2,3]. Since a compound may be able to bind to different targets in different conformations, the algorithm aligns pairs of pharmacophore models sharing the same features which are not necessarily at the exactly same spatial distance. Using the aligned models, a pharmacophore search on a multi-conformation-database is performed to find compounds matching both models. The potentially "dual" ligands are scored by a shape-based comparison with the known active molecules using ShaEP [4]. Using this approach, we performed a prospective fragment-based virtual screening for dual 5-LO/sEH inhibitors. Both enzymes play an important role in the arachidonic acid cascade and are involved in inflammatory processes, pain, cardiovascular diseases and allergic reactions [5,6]. Beside several new selective inhibitors we were able to find a compound inhibiting both enzymes in low micromolar concentrations. The results indicate that the idea of aligned pharmacophore models can be successfully employed for the discovery of dual-target ligands.

References

1. Morphy R, Rankovic Z: *J Med Chem* 2005, **48**:6523-6543.
2. Kabsch W: **.** *Acta Crystallog., Sect A: Found Crystallogr* 1976, **32**:922-923.
3. Kabsch W: **.** *Acta Crystallogr, Sect A: Found Crystallogr* 1978, **34**:827-828.
4. Vainio MJ, Puranen JS, Johnson MS: *J Chem Inf Model* 2009, **49**:492-502.
5. Werz O, Steinhilber D: *Pharmacol Ther* 2006, **112**:701-718.
6. Inceoglu B, Schmelzer KR, Morisseau C, Jinks SL, Hammock BD: *Prostaglandins Other Lipid Mediators* 2010, **82**:42-49.

O12

Text-based similarity searching for hit- and lead-candidate identification

Volker Hähnke
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Department of Health and Human Services, 8600 Rockville Pike, Bethesda, MD 20894, USA
E-mail: volker.hahnke@nih.gov
Journal of Cheminformatics 2012, **4(Suppl 1):**O12

The Pharmacophore Alignment Search Tool (PhAST) is a string-based approach to virtual screening. Molecules are represented by linear sequences which describe their respective pattern of interaction possibilities. The problem of molecule linearization is tackled by applying Minimum Volume Embedding in combination with a Diffusion Kernel to the molecular graph [1,2]. Linear representations are compared using global pairwise sequence alignment [3]. PhAST exhibited enrichment capabilities comparable or superior to most common virtual screening approaches. Compound rankings were proven to be dissimilar to those of other virtual screening techniques. It was shown that emphasis on key interactions through the application of position specific weights in the alignment process significantly increases enrichment.

Significance of chemical similarity was determined in form of p-values of global alignment scores, calculated in an approach that was adapted from its original application to local sequence alignments of protein sequences utilizing Marcov chain Monte Carlo simulation [4]. Bonferroni correction was used to correct p-values with respect to the size of the screening library [5].

PhAST was employed in two prospective applications: A screening for non-nucleoside analogue inhibitors of bacterial thymidine kinase yielded a hit with a distinct structural framework but only weak activity. Screenings for drugs that are not members of the NSAID (non-steroidal anti-inflammatory drug) class as modulators of gamma secretase resulted in a potent modulator with clear structural distinction from the reference compound.

References

1. Shaw R, Jebara T: **Minimum Volume Embedding.** *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics: 21-24 March 2007; San Juan (Puerto Rico)* Omnipress: Meila M, Shen X 2007, 460-467.
2. Smola AJ, Kondor RI: **Kernels and Regularization on Graphs.** *Proceedings of the 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop: 24-27 August 2003; Washington DC* Springer: Schölkopf B, Warmuth, M 2003, 144-158.
3. Durbin R, Eddy S, Krogh A, Mitchison G: **Alignment with affine gap scores.** *Biological Sequence Analysis* Cambridge University Press 1998, 29-31.
4. Hartmann AK: **Sampling rare events: statistics of local sequence alignments.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2002, **65(5 Pt 2)**:056102.
5. Bonferroni CE: **Teoria statistica delle classi e calcolo delle probabilità.** *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936, **8**:3-62.

O13

Understanding nanostructure formation from first principles

Eva Rauls
Universität Paderborn - Theoretische Physik, N3.315, Pohlweg 55, 33098 Paderborn, Germany
E-mail: eva.rauls@uni-paderborn.de
Journal of Cheminformatics 2012, **4(Suppl 1):**O13

In recent years, nanotechnology found its way into various fields of our life, ranging from applications in pharmacy or food technology to miniaturized electronic or optical devices.
Functional surface coatings are among the most well known applications.
Bottom up design is one of the most promising possibilities to build new materials with specifically tailored properties. During the last decade, intense research has been carried out in this area, and many applications have already been put into practice. However, in the world of nanostructures, there are still many effects just waiting to be explored and utilized for new applications.
In computational materials science, we approach the field from the theoretical side in order to gain a deeper and fundamental understanding of structure formation and how the involved processes can efficiently be improved.
In my talk, I will give an overview of some of our recent projects in this field. In the first part, I will discuss the self-organisation driven formation of a covalently bonded molecular network and the catalytic role the surface plays for the reaction between two organic molecules. Imidization reactions - well known in chemistry and biology for a long time, but rather newly discovered by surface scientists - constitute an outstanding tool for the design of nanostructures.
In the second part of my talk, I will show some first results we obtained for the physical properties of metallo-porphyrines adsorbed on a Au-surface and conclude with a brief overview of some of our other fields of research.

O14

Solvent-screening and co-crystal screening for drug development with COSMO-RS

A Klamt
COSMOlogic GmbH&CoKG, Leverkusen and Inst. of Phys. and Theoretical Chemistry, University of Regensburg, Germany
E-mail: klamt@cosmologic.de
Journal of Cheminformatics 2012, **4(Suppl 1):**O14

Bringing active pharmaceutical or agrochemical ingredients (APIs) in solution often is the most demanding step in pharmaceutical and agrochemical development. The COSMO-RS method, which has been originally developed by the author during his 12 years at Bayer, is a unique combination of quantum chemical information and liquid phase thermodynamics and currently is proven to be the most accurate method

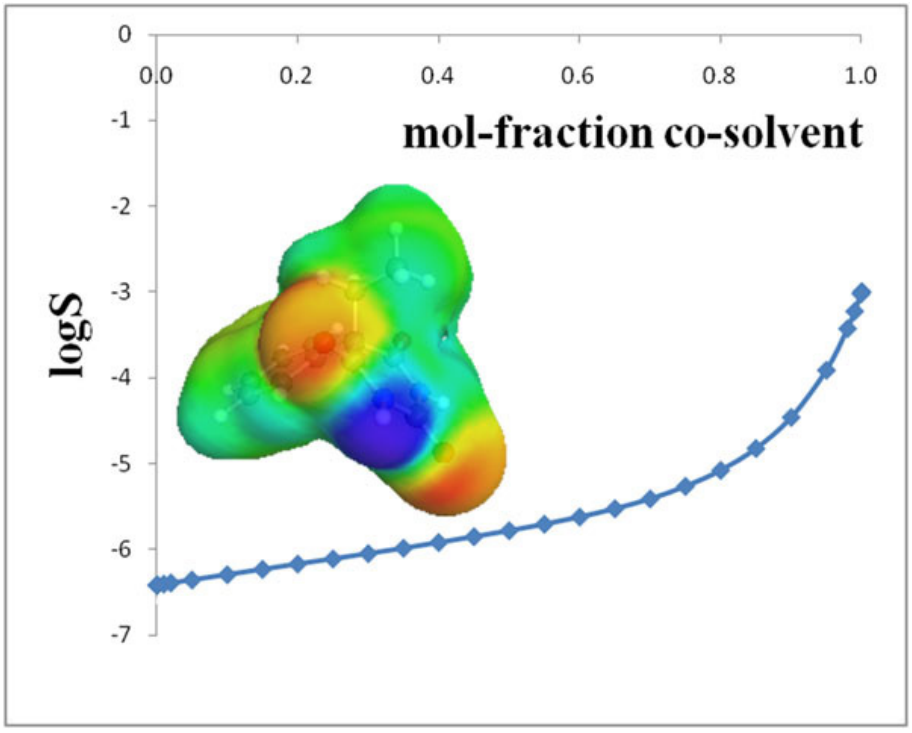

Figure 1(abstract O14)

for predicting the free energy of molecules in solution. Based on COSMO-RS theory the COSMOtherm suite of software tools is able to address a broad range of important aspects of solubilization and thus is an ideally suited toolset for rational solubilization development:
- Solvent screening, including mixtures and variable temperatures
- logP, logD and pK_a prediction, general multi-phase distribution
- conformational preference and tautomer trends in solution
- Co-crystal screening based on mixing enthalpy
- solubility in micellar systems
- solvent-dependent free energy of crystal faces

O15

Assessment of a variety of dispersion-corrected density functional theory calculations used in molecular crystal structure prediction

Bernd Doser*, Marcus A Neumann
Avant-garde Materials Simulation, Merzhauser Straße 177, 79100 Freiburg im Breisgau, Germany
E-mail: doser@avmatsim.de
Journal of Cheminformatics 2012, **4(Suppl 1):**O15

For a long time, the prediction of molecular crystal structures from first principles has been one of the most challenging computational problems. Because of the large number of degrees of freedom, methods with low computational effort have to be used. The price for this efficiency is, however, a loss of accuracy. In recent years, important progress has been made in the development of dispersion corrections [1-3], which raise the accuracy of density functional theory (DFT) calculations to a much higher level.
Based on these developments, the software package GRACE [4] was the first program ever to correctly predict all four crystal structures at the 4th crystal structure prediction blind test, organized by G. M. Day and the Cambridge Crystallographic Data Centre in 2007 [5]. GRACE uses dispersion-corrected density functional theory calculations (DFT-D) to generate reference data to which a tailor-made force field (TMFF) is fitted for each respective molecule [6]. Crystal structures are generated with a Monte-Carlo parallel-tempering algorithm that uses the TMFF for the evaluation of lattice energies. Only the most stable crystal structures according to the TMFF are then re-optimized and re-ranked using DFT-D. The DFT calculations are performed using an interface to the two common ab-inito programs VASP [7] and QuantumESPRESSO [8].
In the most recent, 5th blind test, organized in 2010, the excellent agreement between experimental and predicted structures for small, neutral molecules was confirmed. However, it turned out that the crystal structure prediction of molecular salts and hydrates is still a challenging task. The inability of DFT without exact exchange to describe anions properly, as recently published by Jensen [9], suggests that the inclusion of the Hartree-Fock exchange should improve the accuracy for molecular salts and hydrates. Therefore, we will present an assessment of a variety of current density functionals, in which the dispersion-correction parameters were fitted with respect to a minimization of the deviation between experimental and computed crystal structures.

References

1. Grimme S: *J Comput Chem* 2004, **25**:1463.
2. Neumann MA, Perrin M-A: *J Phys Chem B* 2005, **109**:15531.
3. Grimme S, *et al*: *J Chem Phys* 2010, **132**:154104.
4. [http://www.avmatsim.de].
5. Day GM, *et al*: *Acta Crystallogr B* 2009, **65**:107.
6. Neumann MA: *J Phys Chem B* 2008, **112**:9810.
7. Kresse G, Hafner J: *Phys Rev B* 1993, **47**:558.
8. Giannozzi P, *et al*: *J Phys Condens Matter* 2009, **21**:395502.
9. Jensen F: *J Chem Theory Comput* 2010, **6**:2726.

O16

Semantics vs. statistics in chemical markup

Colin Batchelor
Royal Society of Chemistry, Thomas Graham House, Cambridge, CB4 0WF, UK
E-mail: batchelorc@rsc.org
Journal of Cheminformatics 2012, **4(Suppl 1):**O16

Since the late 1990s, natural language processing (NLP) has seen a massive shift from high-precision, low-recall systems based on small sets of hand-written rules, to methods based on the statistical analysis of large corpora. The field of chemoinformatics, likewise, is dominated by statistical and machine-learning approaches. In recent years, however, pharmaceutical companies have been engaging more and more with Semantic Web technologies, which are largely built around the sorts of hand-written systems that NLP has moved away from this century. We discuss where our current text analysis and Semantic Web efforts at the Royal Society of Chemistry are headed and how we're making use of the unreasonable effectiveness of data.

O17

ChemProspector and generic structures: advanced mining and searching of chemical content

Valentina Eigner-Pitto*, Josef Eiblmaier, Hans Kraut, Larisa Isenko, Heinz Saller, Peter Loew
InfoChem GmbH, München, 81241, Germany
E-mail: ve@infochem.de
Journal of Cheminformatics 2012, **4(Suppl 1):**O17

Chemical information mining has turned into a well-established scientific area over the last five years. Several software solutions exist that are able to identify and extract names of chemical compounds in text documents and convert them into chemical structure-searchable information. Likewise, several programs exist which recognize chemical structures from images and translate them into the computer-readable format, the connection table. However, a still unsolved issue is the automatic abstraction of generic compounds (Markush structures). These usually consist of a core structure image and variable groups specified in the text, in additional images or in tables.
This presentation describes our hybrid approach to extract generic structure information from documents by using combining information science, cheminformatics, computational linguistics and pattern recognition techniques. Experiences with the envisaged methodology and the first results are presented.
This research project is funded by the German Ministry of Economics and Technology. It is part of the THESEUS research programme which has the goal to facilitate access to information, combine data to form new kinds of knowledge and lay the groundwork for new services on the Internet.

O18

Recent and current developments in handling Markush structures from chemical patents

John M Barnard[1*], Geoff M Downs[2]
[1]Digital Chemistry Ltd., Sheffield S26 1HL, UK; [2]Digital Chemistry Ltd., Sheffield S26 1HL, UK
E-mail: john.barnard@digitalchemistry.co.uk
Journal of Cheminformatics 2012, **4(Suppl 1):**O18

The commercially-available database systems for storing and searching Markush structures from chemical patents have undergone little change since their launch some twenty years ago. However, the past few years have seen the area become an active one again for research and development. This presentation offers an overview and commentary on recent and current activity, and discusses the prospects for improved access to structural information in the patent literature [1].
The existing curated Markush databases remain the gold standard, though several groups, both academic and commercial, continue to work on automatic analysis of full-text patents. This has involved not only the identification of specific-structure nomenclature and its conversion to structure-searchable records, but also the attempted reconstruction of searchable representations of complete Markush structures. The advantages and disadvantages of these approaches, and the prospects for their successful commercial exploitation, are discussed.
New commercial software for searching Markush structure databases is being developed by several groups. These employ both conventional substructure search approaches (e.g. ChemAxon, Digital Chemistry), and novel algorithms, in some cases based on various forms of similarity and approximate structure matching (e.g. DecrIPt, IBM). These approaches are summarised and compared, and the opportunity their in-house

implementation provides for integration of chemical patent information with the drug-discovery process is discussed.
Practitioners have long been aware of the inadequacies and complexities of the existing systems, and the extent to which a new generation of systems may satisfy their requirements is discussed. The possible role of systematic evaluation of retrieval performance (in particular, the TREC-CHEM project) is addressed.

Reference

1. Downs GM, Barnard JM: **Chemical Patent Information Systems.** *WIREs Comput Mol Sci* 2011, **1**:727-741, doi:10.1002/wcms.41.

O19

Computational chemistry in pharmaceutical research – where do we stand after 25 years?

Herbert Köppen
Schützenpfad 27, 55218 Ingelheim/D, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**O19

In the early 80th of the last century molecular modeling was introduced as a tool to aid rational drug design in pharmaceutical industry. However in these early days modeling had narrow limits due to the lack of adequate compute power and sophisticated methods. There was a large gap between expectations of medicinal chemists and the power of theoretical methods. Today computational chemistry is an integral part of pharmaceutical research. What has changed over the last decades, what was the driving force of the progress?
The talk is based on the author's experience at Boehringer Ingelheim and will focus on computational chemistry methods with proven value in industrial pharmaceutical research. Amongst others the synergy of virtual screening and combinatorial chemistry and the impact of the recently published structures of GPCRs on pharmaceutical research will be illustrated. Further current challenges, e.g. the consideration of protein flexibility, will be discussed.

O20

Ligand based lead generation - considering chemical accessibility in rescaffolding approaches via BROOD

Li-hsing Wang*, Andreas Evers, Peter Monecke, Thorsten Naumann
Sanofi-Aventis Deutschland GmbH, Industriepark Höchst / G838, D-65926 Frankfurt, Germany
E-mail: Li-hsing.Wang@sanofi.com
Journal of Cheminformatics 2012, **4(Suppl 1):**O20

In pharmaceutical industry ligand based approaches like scaffold hopping, scaffold decoration and me-too approaches, are used to generate lead structures in discovery projects. We use several tools to generate novel lead structures, such as BROOD [1]. BROOD is a software tool which explores chemical space around query molecules based on shape similarity and electrostatics, and it generates analogs of a reference molecule by replacing a selected moiety with fragments from *in silico* fragment databases. The content of these fragment databases has an essential influence on the molecules generated by BROOD. Due to the amount of resources required to synthesize novel chemical compounds, an easy access to chemical compounds is crucial for the broad applicability and for the acceptance of *in silico* approaches which propose novel molecules to the synthetic chemist. In order to consider synthetic accessibility of *in silico* generated molecules, we use our inhouse libraries of drug-like and chemically feasible molecules for the *in silico* fragment generation. In addition, we implemented fragmentation rules which reflect the (retro-)synthetic access to these molecules. The combination of (retro-)synthesis rules and fragments of existing compounds leads mainly to synthetically accessible compound proposals. To identify relevant available compounds, we implemented an approach where *in silico* generated molecules are used as search queries to search in inhouse available compound libraries, e.g. via ROCS [2]. This approach leads directly to internally existing compounds which can be ordered for experimental testing. However, if novel chemical matter is desired, chemical synthesis is necessary. Therefore, as a further extension of our approach, we started from available chemical reagents as input for our *in silico* fragment databases. The available reagents are detected by well-defined chemical reactions, converted to fragments and stored in the *in silico* fragment databases. As the chemical reactions are considered during the virtual synthesis step, the synthetic accessibility of compound proposals is increased. We present the concept of these approaches, examples and typical applications for different targets.

References

1. BROOD, version 1.1.2 and 2.0.0, OpenEye Scientific Software, Inc., Santa Fe, NM, USA. 2010 [http://www.eyesopen.com].
2. ROCS, version 3.1.2, OpenEye Scientific Software, Inc., Santa Fe, NM, USA. 2010 [http://www.eyesopen.com].

O21

Blocking protein-protein interactions: the identification of repetitive turn structures as basis for inhibitor building blocks

Oliver Koch[1,2]
[1]Intervet Innovation GmbH, Schwabenheim, Germany; [2]MOLISA GmbH, Magdeburg, Germany
E-mail: oliver.koch@msd.de
Journal of Cheminformatics 2012, **4(Suppl 1):**O21

Mimetics of secondary structure elements are one promising approach in the design of protein-protein interaction inhibitors, since secondary structure elements are very important recognition motifs in protein-protein interfaces. In helices and turns, the protein backbone provides a scaffold to present the sidechains in the correct orientation for the three-dimensional interaction motif. For both, scaffolds are known that resemble these backbone conformations and can be decorated with sidechains in the right position for mimicking the interaction motif [1]. Benzodiazepines are one example for a successful mimetic of β-turn structures.
However, identifying small chemical scaffolds that mimic turn structures is rather complicated. Turns are irregular structures with a wider variety of possible backbone conformations [2] and for each group of conformations a different scaffold is needed. Furthermore, turn structures are generally not included in analysis of protein-protein interfaces. Due to a lack of information in publicly available databases, regions of the protein chains that are outside helices and β-sheets are generally considered as non-regular structural elements. These non-regular structural elements in proteins are by now almost completely classified as turn structures and available via Secbase for data mining approaches [3].
The results of an exhaustive analysis of turn structures involved in protein-protein interfaces will be presented and the impact on the design of secondary structure element mimetics will be discussed. This is of particular interest since the secondary structure space of protein-protein interfaces is limited and similar interfaces with respect to secondary structure elements exists within proteins showing different overall folds and function [4]. The identification of repetitive turn structures is therefore a valuable approach to predict polypharmacology or identify backbone conformations that could easily be replaced by mimetic building blocks [4]. The decoration of these building blocks with the needed functional sidechain is a good starting point for protein-protein interface inhibitors.

References

1. Yin H, Hamilton AD: **Strategies for targeting protein-protein interactions with synthetic agents.** *Angew Chem Int Ed Engl* 2005, **44**:4130-4163.
2. Koch O, Klebe G: **Turns revisited: a uniform and comprehensive classification of normal, open, and reverse turn families minimizing unassigned random chain portions.** *Proteins* 2009, **74**:353-367.
3. Koch O, Cole J, Block P, Klebe G: **Secbase: database module to retrieve secondary structure elements with ligand binding motifs.** *J Chem Inf Model* 2009, **49**:2388-2402.
4. Koch O: **The use of secondary structure element information in drug design: polypharmacology and conserved motifs in protein-ligand binding and protein-protein interfaces.** *Future Med Chem* 2011, **3**:699-708.

O22

Improvements in docking scoring functions: the physics-based perspective

Xavier Barril
Dept. Fisicoquímica, Facultat de Farmàcia, Av. Joan XXIII, s/n 08028 - Barcelona, Spain
E-mail: xbarril@ub.edu
Journal of Cheminformatics 2012, **4(Suppl 1):**O22

Docking is possibly the most widely used technique in structure-based drug design. It is recognized as extremely useful to guide the development of active molecules as well as in virtual screening to identify new leads. But docking is also notoriously imperfect and many aspects need to improve. In this talk I will present our work on two different sources of errors in the docking predictions: protein-ligand interaction potentials and internal energy of the ligands. The use of higher-level calculations offers an opportunity to obtain better quality results and offers a glimpse of the full potential of the technique, but introducing this information back into fast and general scoring functions remains a challenge.

O23

Molecular modeling of lipid drug formulations

Woldeamanuel A Birru*, Dallas B Warren, Christopher JH Porter, Colin W Pouton, David K Chalmers
Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Pde, Parkville, VIC, Australia
E-mail: emanuel.birru@monash.edu
Journal of Cheminformatics 2012, **4(Suppl 1):**O23

Lipid formulations can improve the bioavailability of drugs that have low aqueous solubility. A variety of chemical compounds, including triglyceride oils (lipids), fatty acid esters and surfactants, can be included in lipid formulations. This heterogeneity makes spectroscopic study of the in ternal structure of formulation difficult. Understanding of lipid formulations at a molecular level will greatly improve our knowledge of in vivo dispersion and solubilisation patterns of lipid formulations.
Molecular dynamics studies have provided useful insight into the structure and dynamics of different types of aggregates, including mixed glycerides with and without propylene glycol [1] and bile salts [2]. To date, such studies have not been performed on lipid drug formulations. The objective of this research is to develop a molecular dynamics protocol to examine the interaction between drugs and formulations at the atomic level. To evaluate and parameterize the force field of choice we are calculating Gibbs free energy of solvation of a number of alcohols and short poly -(ethylene glycol) polymers. Following this, the aggregation behaviour of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), sodium glycochenodeoxycholate (GDX), different digestion products and polyethylene glycol surfactants will be investigated. Moreover, the phase diagram s of three component systems composed of i) bile salts, digested products and water and ii) surfactants, lipids and water will be modelled.
Simulations are being performed using the molecular dynamics software suite GROMACS. Calculations are being performed on a high performance computing cluster at the Victorian Life Sciences Com putation Initiative (VLSCI). The methods highlighted in this study will prove to be an essential tool for formulators of lipid systems for oral administration.

References

1. Warren DB, Chalmers DK, Pouton CW: Structure and Dynamics of Glyceride Lipid Formulations, with Propylene Glycol and Water. *Mol Pharm* 2009, **6**:604-614.

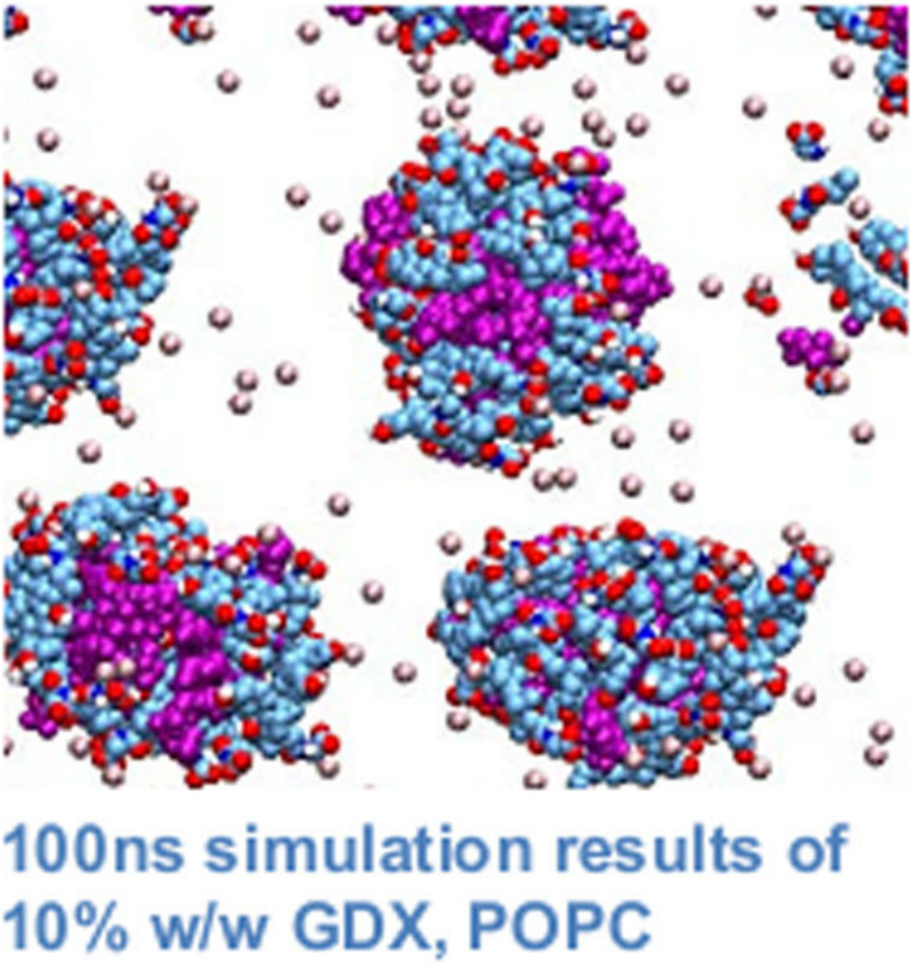

Figure 1(abstract O23)

2. Warren DB, *et al*: **Molecular Dynamics Simulations of Spontaneous Bile Salt Aggregation.** *Colloid Surface A* 2006, **280**:182-193.

O24

Virtual screening for plant PARP inhibitors – what can be learned from human PARP inhibitors?

Peter-Paul Heym[1*], Wolfgang Brandt[1], Ludger A Wessjohann[1,2], Hans-Joachim Niclas[2]
[1]Leibniz-Institute for Plant Biochemistry, Halle, 06114, Germany; [2]Agrochemical Institute Piesteritz e.V. (AIP), Wittenberg, 06886, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**O24

The functions of Poly(ADP-ribose) polymerase enzymes (PARPs) in general are best studied based on human PARP-1 (*Hs*PARP-1). *Hs*PARP-1 is well investigated because pharmacological modulation of its activity modulates DNA-binding of antitumor drugs [1]. In contrast to human PARP enzymes, the exact role of PARPs in plants remains to be elucidated. Different stresses activate plant PARP enzymes to mediate DNA repair and (programmed) cell death whereas the addition of PARP inhibitors decreases the degree of cell death [2]. Therefore, the development of plant PARP inhibitors might be a way to increase the tolerance against abiotic stress.

Initial to searches in commercial databases for potential plant PARP inhibitors, a virtual screening route had to be established for human PARP-1 inhibitors. Simultaneously, every step in that procedure was applied on a plant PARP enzyme to investigate the differences of both active sites. All differences have been evaluated statistically, e.g. using receiver-operator characteristics (ROC) and power analyses. At the end of that parallel screening route, a docking threshold for *Arabidopsis thaliana* L. PARP-1 (*At*PARP-1) could be derived by knowledge transfer from the corresponding human receptor and its inhibitors.

Knowing the differences of the human and plant screening routes, predictions of the applicability of that multi-step process on a commercial database have been explored. Finally, the developed virtual screening route has been applied to screen a commercial database for *At*PARP-1 inhibitors. From 20 compounds tested so far *in vitro*, 13 show inhibitory effects.

References

1. Ferraris D: **Evolution of Poly(ADP-ribose) Polymerase-1 (PARP-1) Inhibitors. From Concept to Clinic.** *J Med Chem* 2010, **53**:4561-4584.
2. DeBlock M: **Poly(ADP-ribose) polymerase in plants affects energy homeostasis, cell death and stress tolerance.** *The Plant Journal* 2005, **41**:95-106.

POSTER PRESENTATIONS

P1

Virtual screening and in silico design of novel inhibitors of bacterial lectins

Jan Alán[1,2*], Petr Kulhánek[1,2], Jaroslav Koča[1,2]
[1]CEITEC – Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic; [2]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic
E-mail: alan@chemi.muni.cz
Journal of Cheminformatics 2012, **4(Suppl 1):**P1

Bacterium *Pseudomonas aeruginosa* is a human opportunistic pathogen. It can cause infection of immunocompromised people or people suffering from cystic fibrosis, which is often fatal. Bacterial colonization of human tissues is mediated by interaction of bacterial surface proteins – lectins – with cellular surface carbohydrates. PA-IIL is *Pseudomonas aeruginosa* tetrameric lectin, which contains two calcium ions in each binding site and recognizes fucosylated oligosaccharides [1]. Saturation of bacterial surface lectins by specially designed compounds might prevent adhesion to host tissues and thus suppress the infection. In this work, virtual screening and docking were used for identification of compounds that might inhibit this interaction and be potentially used as a new generation of antibiotics.

Two different approaches for identification of promising compounds were employed. Subset of drug-like molecules was docked into the PA-IIL binding site by Autodock Vina [2] and Dock 6 [3]. Interesting ligands were then selected by identifying those with highest score provided by both programs. In the second approach, we focused on the identification of interesting molecular fragments, which should be attached to already pre-docked carbohydrate. The carbohydrate serves as a targeting agent and newly identified fragments increase its interaction with the lectin. In both approaches, ZINC library [4] was the source of ligands and fragments.

Acknowledgements: The work is supported by the Ministry of Education (MSM0021622413, LC6030, ME08008), the Czech Science Foundation (GA/303/09/1168, GD301/09/H004) of the Czech Republic, the European Community's Seventh Framework Programme (grant agreement n° 205872) and the project "CEITEC - Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068) from European Regional Development Fund.

References

1. Adam J, *et al*: **Engineering of PA-IIL lectin from *Pseudomonas aeruginosa* - Unravelling the role of the specificity loop for sugar preference.** *BMC Struct Biol* 2007, **7**:36.
2. Trott O, *et al*: **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.** *J Comput Chem* 2010, **31**:455-461.
3. Kuntz ID, *et al*: **A geometric approach to macromolecule-ligand interactions.** *J Mol Biol* 1982, **161**:269-288.
4. Irwin JJ, *et al*: **ZINC–a free database of commercially available compounds for virtual screening.** *J Chem Inf Model* 2005, **45**:177-182.

P2

DecoyFinder, a tool for finding decoy molecules

Cereto Massagué Adrià[*], S Garcia-Vallvé, G Pujadas
Tarragona, Spain
Journal of Cheminformatics 2012, **4(Suppl 1):**P2

DecoyFinder is a graphical tool which helps finding sets of decoy molecules for a given group of active ligands. It does so by finding molecules which have a similar number of rotational bonds, hydrogen bond acceptors, hydrogen bond donors, logP value and molecular weight, but are chemically different, which is defined by a maximum Tanimoto value threshold between active ligand and decoy molecule MACCS fingerprints. Optionally, a maximum Tanimoto value threshold can be set between decoys in order assure chemical diversity in the decoy set.

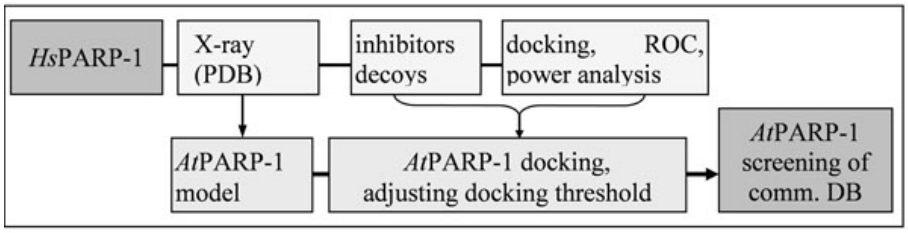

Figure 1(abstract O24) Key steps in virtual screening routes for human and *Arabidopsis thaliana* L. PARP-1. The results have been used to successfully apply the screening process for *At*PARP-1 on a commercial database.

During the talk, the algorithm used by DecoyFinder in order to look for decoys sets will be described in detail and some examples of its application will be described and discussed.

P3

Targeting protein-protein interactions using methods of cheminformatics

Art Cherkasov
Department of Urological Sciences, University of British Columbia, Vancouver, BC, V6H 3Z6, Canada
E-mail: artc@interchange.ubc.ca
Journal of Cheminformatics 2012, **4(Suppl 1):**P3

We have recently mapped the protein interaction networks of methicillin-resistant Staphylococcus aureus that revealed its scale-free organization with characteristic presence of highly-connected hub proteins that are critical for bacterial survival [1]. Here we report the discovery of highly selective nanomolar inhibitors for one such hub target - staphylococcal pyruvate kinase. The lead compound has been identified through synergetic combination of methods of high-throughput screening and cheminformatics; its further synthetic modifications resulted in much improved antimicrobial properties. Further lead optimization yielded drug candidates with picomolar activity against methicillin-resistant Staphylococcus aureus.
Considering a notable lack of recent reports on novel antibacterial targets and cognate antibacterial compounds, this study provides a valuable perspective on the development of a new generation of antimicrobials. Equally noteworthy, the results of the current work highlight the importance of cheminformatics-driven exploration of chemical space around initial high throughput screening hits.

Reference

1. Cherkasov A, Hsing M, Zoraghi R, Foster LJ, See RH, Stoynov N, Jiang J, Kaur S, Lian T, Jackson L, Gong H, Swayze R, Amandoron E, Hormozdiari F, Dao P, Sahinalp C, Santos-Filho O, Axerio-Cilies P, Byler K, McMaster WR, Brunham RC, Finlay BB, Reiner NE: *J Proteome Res* 2011, **10**:1139-1150.

P4

Design of multi-target activity landscapes that capture hierarchical activity cliff distributions

Dilyana Dimova*, Jürgen Bajorath
Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dahlmannstr. 2, D-53113 Bonn, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P4

For compounds with activity against multiple targets the ensuing multi-target structure-activity relationships (mtSARs) are generally difficult to analyze. However, understanding mtSARs is often of critical importance for compound design and optimization [1]. In addition, the detection and analysis of activity cliffs also plays a crucial role in comprehensive SAR exploration [2,3]. Activity cliffs represent the most prominent feature of activity landscapes, which can be graphically represented by models that integrate molecular similarity and potency relationships [2,4]. Different activity landscape representations have been introduced. These activity landscape designs have in common that they all focus on activity against a single or at most two biological targets (the latter case giving rise to selectivity landscapes [3]). For compounds active against multiple targets, landscape representations cannot be obtained directly on the basis of currently available models and new design concepts are required. Here we introduce a methodology to derive and visualize multi-target activity landscapes and systematically analyze activity cliff distributions [5]. The framework is based on a general activity cliff classification scheme. Multi-target activity landscapes are visualized as graphs where nodes represent individual compounds and edges activity cliffs. In addition, node proximity indicates molecular similarity. The methodology has been applied to derive landscape models for various compound data sets with activity against multiple targets belonging to different families. The resulting representations identify single and multi-target activity cliffs and reveal hierarchical cliff distributions. From these landscape models, compounds forming complex activity cliffs can be readily selected.

References

1. Kubinyi H: **Similarity and Dissimilarity. A Medicinal Chemist's View.** *Perspect Drug Discovery Des* 1998, **9-11**:225-252.
2. Maggiora GM: **On Outliers and Activity Cliffs - Why QSAR Often Dissapoints.** *J Chem Inf Model* 2006, **46**:1535.
3. Bajorath J, Peltason L, Wawer M, Guha R, Lajiness MS, Van Drie JH: **Navigating Structure-Activity Landscapes.** *Drug Discovery Today* 2009, **14**:698-705.
4. Wassermann AM, Wawer M, Bajorath J: **Activity Landscape Representations for Structure-Activity Relationship Analysis.** *J Med Chem* 2010, **53**:8209-8223.
5. Dimova D, Wawer M, Wasserman AM, Bajorath J: **A Design of Multitarget Activity Landscapes That Capture Hierarchical Activity Cliff Distributions.** *J Chem Inf Model* 2011, **51**:258-266.

P5

Structured chemical class definitions and automated matching for chemical ontology evolution

Lian Duan[1,2*], Janna Hastings[1,3], Paula de Matos[1], Marcus Ennis[1], Christoph Steinbeck[1]
[1]European Bioinformatics Institute, Cambridge, UK; [2]ETH, Zurich, Switzerland; [3]University of Geneva, Switzerland
E-mail: dlian@ebi.ac.uk
Journal of Cheminformatics 2012, **4(Suppl 1):**P5

Ontologies encode the knowledge of human experts in order to allow computers to automate common tasks in a domain. They are hierarchically organised and backed by computational logic which allows automated inferences of the implicit consequences of explicitly stated knowledge. ChEBI is a database and ontology of chemical entities of biological interest [1]. Within the ontology, chemical entities are classified based on shared structural features and also based on their roles and activities in biological systems. For example, the chemical class 'aminopyridine' is defined as 'Compounds containing a pyridine skeleton substituted by one or more amine groups', while an example of a role based class is 'antiviral drug', which groups together chemical entities that are used as antiviral drugs, regardless of their chemical structure. We have developed a novel semi-automated system for creating structure-based chemical class definitions. Our tool allows curators to draw and visually define shared structural features for classes of chemicals, which definitions are then used to automatically detect class membership across the full chemical database. The front end is based on an extended JChemPaint [2] and the Google Web Toolkit, and the back-end on a custom extension of the Chemistry Development Kit [3]. With this tool, it is possible to define chemical classes based on molecular skeletons, substitute groups, arbitrary parts including cycles of arbitrary length, formulae and overall properties, and these features can be combined using nested logical operators. Matching these definitions to candidate structures from the database is accomplished by means of an in-memory matching procedure, validated against the existing manually curated classification in ChEBI, allowing us to iteratively refine both the definitions of classes as well as to evolve the quality of the classification in ChEBI.

References

1. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M: **ChEBI: a database and ontology for chemical entities of biological interest.** *Nucl Acids Res* 2008, **36(Suppl. 1)**:D344-D350.
2. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E: **The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics.** *J Chem Inf Comput Sci* 2003, **43**:493-500.
3. Krause S, Willighagen E, Steinbeck C: **JChemPaint- Using the Collaborative Forces of the Internet to Develop a Free Editor for 2D Chemical Structures.** *Molecules* 2000, **5(1)**:93-98.

P6

Structure and thermodynamics of nonaqueous solvation by integral equation theory

Roland Frach*, Jochen Heil, Stefan M Kast
Technische Universität Dortmund, Dortmund, 44227, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P6

Electronic structure theory under the influence of apolar solvents suffers from substantial methodical difficulties since in this case the solvent-induced solute polarization originates mainly from specific directional interactions and higher electric multipoles. Continuum solvation models based on the dielectric solvent response such as the PCM approach ignore such interactions and can therefore not adequately model solvation effects in nonaqueous environments.

The "embedded cluster reference interaction site model" (EC-RISM) [1] retains the granularity of the solvent and represents a microscopically more detailed and therefore improved approach towards solvation modeling. EC-RISM is based on a self-consistent solution of solvent distribution functions described by a 3D integral equation theory and solute electronic structure by mapping the solvent charge distribution onto discrete, solute-embedding point charges. In aqueous solution EC-RISM theory is capable of calculating pK_a shifts [1] and tautomer ratios relatively fast and with high accuracy [2].

Here we outline the strength of the integral equation model by studying benzene and hexafluorobenzene solutions. In particular, the thermodynamics of differential solvation is quantified for organic compounds dissolved in these media. Moreover, it is shown that the respective solvent structures around particular solutes differ strongly, possibly leading to changes in the thermodynamic stability scale of various isomers which are not reproduced by the PCM model.

References

1. Kloss T, Heil J, Kast SM: **Quantum chemistry in solution by combining 3D integral equation theory with a cluster embedding approach.** *J Phys Chem B* 2008, **112**:4337-4343.
2. Kast SM, Heil J, Güssregen S, Schmidt KF: **Prediction of tautomer ratios by embedded-cluster integral equation theory.** *J Comput.-Aided Mol Des* 2010, **24**:343-353.

P7

Assessment of a probabilistic framework for combining structure- and ligand-based virtual screening

Simone Fulle*, Stuart M Armstrong, Paul W Finn, Garrett M Morris
InhibOx, Limited, Oxford, UK
E-mail: simone.fulle@inhibox.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P7

A wide variety of structure- and ligand-based virtual screening approaches have been developed that aim at finding potential leads to initiate drug discovery efforts. Since each method has its strengths and weakness, combining the outcome of different structure- and ligand-based approaches can be expected to decrease the number of false positive predictions. However, a reliable fusion of information from different methods is challenging. This holds true in particular for new target structures, where target specific performance experiences are missing.

Here, we assess the performance of a probilistic framework approach [1] that combines structure- and ligand-based information in a meaningful way by assigning probabilities that any two molecules are active. The approach is validated using two popular docking methods (GOLD and AutoDock) and an in-house ligand-based screening approach (ElectroShape [2]). Results of similarity search and docking calculations for the Directory of Useful Decoys (DUD) [3] are combined through rank fusion as well as a probabilistic framework approach.

The study will be used to answer questions such as: How far do the virtual screening-approaches used provide complementary or redundant hit lists? Does the fusion of structure- and ligand-based approaches consistently outperform any single screening metric? Using a probabilistic framework approach, is it possible to obtain a quantification of the confidence that any molecule will be active?

References

1. Swann S, Brown S, Muchmore S, Patel H, Merta P, Locklear J, Hajduk P: **A unified, probabilistic framework for structure- and ligand-based virtual.** *J Med Chem* 2011, **54**:1223-1232.
2. Armstrong MS, Finn PW, Morris GM, Richards WG: **Improving the accuracy of ultrafast ligand-based screening: incorporating lipophilicity into ElectroShape as an extra dimension.** *J Comput Aided Mol Des* 2011, **25**:785-790, doi:10.1007/s10822-011-9463-8.
3. Huang N, Shoichet BK, Irwin JJ: **Benchmarking sets for molecular docking.** *J Med Chem* 2006, **49**:6789-6801.

P8

Applying the unified pH scale: absolute acidities in the gas phase and anchor points for eleven representative liquid media

Sascha K Goll[1*], Daniel Himmel[1], Ivo Leito[2], Ingo Krossing[1]
[1]Institute for Inorganic and Analytical Chemistry, Freiburger Materialforschungszentrum FMF and Freiburg Institute for Advanced Studies (FRIAS), Albert-Ludwigs-Universität Freiburg, Freiburg, 79104, Germany; [2]Institute of Chemistry, University of Tartu, Tartu, 50411, Estonia
E-mail: sascha.goll@ac.uni-freiburg.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P8

The investigations on our recently introduced unified acidity scale [1] based on the absolute chemical potential of the proton pointed out the inadequateness of the established GA scale. Earlier it was inter alia found that, when trying to correlate pK_a with GA values, in several cases the correlation was broken without any sufficient explanation [2,3]. However, the GA does not take into account the pressure dependent speciation in the gas phase. In this contribution we systematically extend the theory of acidity in the gas phase from standard GAs and GBs to the real existing bulk phases [4].

Furthermore we present the *rCCC* (relaxed COSMO cluster-continuum) model [5], a quantum chemical solvation model for the calculation of *Gibbs* solvation energies of the proton with good accuracy. The *rCCC* values can be used to anchor individual pH scales in different solvents to our universal scale.

References

1. Himmel D, Goll SK, Leito I, Krossing I: **A unified pH scale for all phases.** *Angew Chem Int Ed Engl* 2010, **49**:6885-6888.
2. Koppel IA, Koppel J, Pihl V, Leito I, Mishima M, Vlasov VM, Yagupolskii LM, Taft RW: **Comparison of Brønsted acidities of neutral CH acids in gas phase and dimethyl sulfoxide.** *J Chem Soc, Perkin Trans 2* 2000, 1125-1133.
3. Kütt A, Leito I, Kaljurand I, Sooväli L, Vlasov VM, Yagupolskii LM, Koppel IA: **A comprehensive self-consistent spectrophotometric acidity scale of neutral Brønsted acids in acetonitrile.** *J Org Chem* 2006, **71**:2829-2838.
4. Himmel D, Goll SK, Leito I, Krossing I: to be published.
5. Himmel D, Goll SK, Leito I, Krossing I: **Anchor points for the unified Brønsted acidity scale: The rCCC model for the calculation of standard Gibbs energies of proton solvation in eleven representative liquid media.** *Chem Eur J* 2011, **17**:5808-5826.

P9

Mechanistic DFT studies – helicate-type complexes with different alcylic spacers

Verena Gossen*, Gerhard Raabe, Markus Albrecht
Institute of Organic Chemistry, RWTH Aachen University, Landoltweg 1, 52074 Aachen, Germany
E-mail: Verena.Gossen@rwth-aachen.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P9

Metal-controlled self-assembly of complexes is of high interest in the field of Supramolecular Chemistry [1,2]. In the current study, we synthesized binuclear complexes with different spacers and study the influence of chain length on their relative energy. The considered complexes prefer the zigzag conformation. Thus a bridge with an odd number of methylene units forms a meso-Helicate ($\Delta\Lambda$ or $\Lambda\Delta$) and one with an even number leads to a Helicate ($\Delta\Delta$ or $\Lambda\Lambda$) (figure 1) [3,4].

Comparison of the calculated transition energies for the non-dissociative interconversions of the diastereomeres with experimental results provides inside into the isomerization process. Moreover, insertion of different cations (templates) into the cavities of the binuclear complexes and corresponding calculations allow prediction of their influence on the isomerization.

Enlargement of the studied system results in binuclear complexes with imino-bridged ligands. The obtained computational results provide a possible explanation for the experimentally observed high diastereoselectivity.

As the DFT functionals like B3LYP do not describe long-range interactions properly, we chose the coulomb-attenuating method CAM-B3LYP [5] which corrects the exchange interaction at long range. The complexes with Ti(IV) in their helical or meso form have been geometrically optimized at the CAM-B3LYP level of theory with the TZVP basis set and

Bridge	Form	Energy in a.u.
methylene	ΔΔ	–2519.676584
methylene	ΔΛ	–2519.708944
ethylene	ΔΔ	–2637.656670
ethylene	ΔΛ	–2637.637550
propylene	ΔΔ	–2755.551314
propylene	ΔΛ	–2755.570690

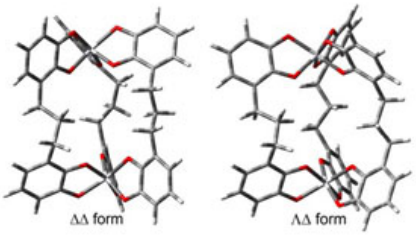

Figure 1(abstract P9) Geometrically optimized complexes with Ti(IV); left: energies of the complexes with different alcylic spacers; right: ΔΔ and ΛΔ form of the complex with a propylene spacer.

MDF10 as ECP for Ti(IV) as implemented in the program package Gaussian09 [6].

References

1. Albrecht M: *Chem Soc Rev* 1998, **27**:281.
2. Albrecht M, Janser I, Fröhlich R: *Chem Commun* 2005, 157.
3. Albrecht M, Kotila S: *Angew Chem* 1996, **108**:1299.
4. Albrecht M, Kotila S: *Angew Chem* 1995, **107**:2285.
5. Yanai T, Tew D, Handy N: *Chem Phys Lett* 2004, **393**:51.
6. **Gaussian 09, Revision A02.** Gaussian Inc., Wallingford CT 2009.

P10

Tabu search based global optimization algorithms for problems in computational chemistry

Christoph Grebner*, Johannes Becker, Daniel Weber, Bernd Engels
Institute of Physical and Theoretical Chemistry, University of Wuerzburg, Emil Fischer Strasse 42, Wuerzburg, 97074, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P10

Efficient searches for global minima of highly dimensional functions with numerous local minima are central for the solution of many problems in computational chemistry. Well known examples are the identification of the most stable conformer of molecules possessing a high number of freely rotatable bonds [1] or the equilibration phase for QM/MM computations. Mathematically, both represent global optimization problems in which the potential energy function of the molecule is the objective function while the coordinates used for representing the structural arrangement of the system are the variables.

Based on an analysis of well-known metaheuristic algorithms, several new global optimization algorithms based on Tabu Search (TS) were developed in our group [2,3], which are using a *steepest descent – modest ascent* strategy. In a first application the Gradient Only Tabu Search (GOTS) was shown to be applicable to conformational search problems [4].

Further test calculations showed its high efficiency in comparison to other global search algorithms like Molecular Dynamics, Simulated Annealing or Monte Carlo Minimization (MCM). The tests also revealed that the efficiency of GOTS can be enhanced dramatically by combining GOTS (searches the nearest neighbourhood highly efficient) with short MCM simulations (samples the phase space more widely) [5]. The new algorithm is implemented into a program for conformational search, providing several different force fields and sampling algorithms.

The investigations also pointed out that a key point is the algorithm for the *modest ascent* strategy. For providing a smoother scan of the potential energy surface as well as a more accurate description of the minimum energy path between two found minima, the Dimer-Method for transition state search was implemented into GOTS-MCM. First tests show that it is much more efficient than previous versions of GOTS. The new algorithm is applied to solvation of biomolecules to provide global optimized solvent shells [6].

In future, the GOTS algorithm will also be used for a proper prediction of reaction pathways by global optimized minimum energy paths and molecular modelling or docking of potential drug molecules to enzymes.

References

1. Schlund S, Schmuck C, Engels B: *Chem Eur J* 2007, **13**:6644-6653.
2. Stepanenko S, Engels B: *J Comp Chem* 2007, **28**:601-611.
3. Stepanenko S, Engels B: *J Comp Chem* 2008, **29**:768-780.
4. Stepanenko S, Engels B: *J Phys Chem A* 2009, **113**:11699-11705.
5. Grebner C, Becker J, Stepanenko S, Engels B: *J Comp Chem* 2011, **32**:2245-2253.
6. Manuscript in preparation.

P11

Synergistic approach of structure-based and ligand-based drug design for the development of selective cannabinod receptor ligands

Robert Günther*, Peter Brust
Institute of Radiopharmacy, Helmholtz-Zentrum Dresden-Rossendorf, Research Site Leipzig, Leipzig, 04318 Germany
E-mail: r.guenther@hzdr.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P11

Cannabinoid receptors (CB) are G-protein coupled receptors involved in many physiological processes, like learning, appetite, nociception and others. Two subtypes (termed CB1 and CB2) are involved in slightly different processes [1]. Thus, it is important to gain more insight into the the cannabinoid receptor system and the potential effects of cannabinoid therapeutics.

By combining [2] 3D-QSAR, pharmacophore modeling, comparative modeling and molecular docking we could identify features responsible for receptor subtype specificity.

Various pharmacophore models were derived from in-house libraries and data available in the literature. 3D structures of both receptor subtypes were created employing comparative modeling methods. The models were subjected to molecular simulations in solvated lipid bilayers to sample different receptor conformations. The models were used for molecular docking studies with small compound libraries. Employing the data obtained in the pharmacophore/3D-QSAR studies as additional constraints delivered valuable information on affinity and selectivity of the compounds towards CB1 and CB2. The results from this synergistic modeling approach could improve our understanding of the protein–ligand interactions involved.

Acknowledgements: This synergistic approach has been implemented into the MOE modeling package (MOE: Chemical Computing Group Inc. Montreal. H3A 2R7 Canada. http://www.chemcomp.com).

Reference

1. Pertwee RG: **Ligands that target cannabinoid receptors in the brain: from THC to anandamide and beyond.** *Addict Biol* 2008, **13**:147-159.

P12

Potency-directed similarity searching using support vector machines

Kathrin Heikamp*, Anne M Wassermann, Jürgen Bajorath
Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Dahlmannstr. 2, D-53113 Bonn, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P12

Searching for active compounds in screening databases is one of the main tasks in chemoinformatics [1,2]. For this purpose, different approaches have been developed, with an increasing interest in machine

learning and data mining methods [3]. Among these, support vector machine (SVM) learning has proven to be a powerful search technique in many instances [3]. Several applications of SVMs have been reported that focus on compound ranking in similarity searching [4-6]. However, similarity search and machine learning methods that are commonly utilized for virtual screening generally do not take compound potency information into account. Regardless of the applied methods, one typically attempts to distinguish "active" from "inactive" compounds. With the exception of QSAR models adapted for compound screening, no approaches have thus far been introduced that incorporate potency information as a parameter and direct search calculations toward the recognition of potent hits. Here, an SVM approach for potency-directed virtual screening is introduced [7]. A newly designed structure-activity kernel and an SVM linear combination model take potency information of reference molecules into account. Applied to high-throughput screening data sets, this potency-directed SVM approach met or exceeded the recall performance of standard SVM ranking and led to a notable enrichment of highly potent hits in database selection sets.

References

1. Stumpfe D, Bajorath J: **Applied virtual screening: strategies, recommendations, and caveats.** *Methods and Principles in Medicinal Chemistry. Virtual Screening. Principles, Challenges, and Practical Guidelines* Weinheim: Wiley-VCH: Sotriffer C 2011, 73-103.
2. Eckert H, Bajorath J: **Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches.** *Drug Discov Today* 2007, **12**:225-233.
3. Geppert H, Vogt M, Bajorath J: **Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation.** *J Chem Inf Model* 2010, **50**:205-216.
4. Jorissen RN, Gilson MK: **Virtual screening of molecular databases using a support vector machine.** *J Chem Inf Model* 2005, **45**:549-561.
5. Wassermann AM, Geppert H, Bajorath J: **Searching for target-selective compounds using different combinations of multiclass support vector machine ranking methods, kernel functions, and fingerprint descriptors.** *J Chem Inf Model* 2009, **49**:582-592.
6. Agarwal S, Dugar D, Sengupta S: **Ranking chemical structures for drug discovery: a new machine learning approach.** *J Chem Inf Model* 2010, **50**:716-731.
7. Wassermann AM, Heikamp K, Bajorath J: **Potency-directed similarity searching using support vector machines.** *Chem Biol Drug Des* 2011, **77**:30-38.

P13

Status of the InChI algorithm and InChI trust

Stephen Heller
Gaithersburg, MD USA
Journal of Cheminformatics 2012, **4(Suppl 1):**P13

Progress of the IUPAC InChI/InChIKey project continues to move ahead quote well. Use of the algorithm has increased over the past year to the point that numerous publications use and refer to InChI. The Markush and Polymer/Mixtures IUPAC InChI working groups have completed their initial work and/or are moving ahead at a good pace. Publicity for the project is good and has resulted in considerable increase in the usage of the InChI algorithm. The Trust web site is now available to the public and is updated regularly. Even CAS now allows for an InChI string to be used as input for a SciFinder search. Extensions to the project's current capabilities are being developed by a number of expert, experienced individuals and groups from various areas of chemistry. This presentation will describe the current technical state of the InChI algorithm and how the InChI Trust is working to assure the continued support and delivery of the InChI algorithm.

P14

A flexible-hydrogen interaction model for protein-ligand docking

Angela M Henzler*, Sascha Urbaczek, Benjamin Schulz, Matthias Rarey
Center for Bioinformatics, University of Hamburg, Hamburg, 20146, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P14

Although some docking methods accounting for protein flexibility exist, most large scale virtual screening approaches work with rigid protein models. A first step towards flexibility integration is the consideration of degrees of freedom resulting from hydrogens, especially, if involved in hydrogen bonding. To account for this type of flexibility, we present a flexible-hydrogen interaction model as part of a descriptor-based docking technique.

The model discretizes interaction spheres of rigid and flexible hydrogen-bond donors and acceptors as interaction spots. A spot has an associated interaction direction which indicates hydrogen or lone pair orientation, and thus, the potential location of a hydrogen-bond counterpart. This new flexible-hydrogen interaction model is combined with a novel approach to describe hydrophobic contacts. Both are introduced in our descriptor-based docking approach named TrixX [1]. TrixX handles ligand flexibility by applying a conformer ensemble approach [2]. The latter allows for the use of efficient indexing techniques upon virtual screening. The discretized, flexible-hydrogen model proposes potential hydrogen and lone pair positions. However, these proposals may still slightly differ from their actual location which can be only determined in presence of pose and active site, i. e., after the docking stage. In order to grant a thorough assessment of hydrogen bonds, thereby, the predicted poses are forwarded to an efficient post-optimization of the hydrogen-bond network. It optimally aligns hydrogens, identifies favorable tautomeric and protonation states, and evaluates the predicted pose [3].

Redocking of the Astex Diverse Set [4] shows that the described docking method produces results in good agreement with co-crystalized ligand structures. Several case studies using different levels of discretization and post-optimization, illustrate the influence of our presented procedures in ligand placement and scoring. The studies highlight the impact of flexible hydrogens and lone pairs during docking and confirm the introduction of our flexible hydrogen model.

References

1. Schlosser J, Rarey M: **Beyond the Virtual Screening Paradigm: Structure-Based Searching for New Lead Compounds.** *J Chem Inf Model* 2009, **49**:800-809.
2. Griewel A, Kayser O, Schlosser J, Rarey M: **Conformational Sampling for Large-Scale Virtual Screening: Accuracy versus Ensemble Size.** *J Chem Inf Model* 2009, **49**:2303-2311.
3. Lippert T, Rarey M: **Fast automated placement of polar hydrogen atoms in protein-ligand complexes.** *J Cheminform* 2009, **1**:13.
4. Hartshorn MJ, *et al*: **Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance.** *J Med Chem* 2007, **50**:726-741.

P15

Revisiting the dataflow principle for chemical information processing

Wolf D Ihlenfeldt
Xemistry GmbH, Königstein, Germany
E-mail: wdi@xemistry.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P15

Dataflow systems, such as Pipeline Pilot or KNIME have become important mainstream tools for data processing in chemistry. These established systems are all implemented relying on a data model emphasizing a strict row/column-centric data table view which does not facilitate interaction with individual chemistry objects, or non-uniform data contents.

Resuming our pioneering work which resulted in the implementation of the first dataflow system for chemistry [1], we present in this contribution a different, object-centric approach for the design of re-usable chemical information processing sequences. Our system is based on the metaphor of a factory floor, instead of opaque pipelines. Individual machining stations perform configurable processing steps on objects such as structures, reactions, datasets or tables. Objects are transported between these - or temporarily set on the factory floor for storage or inspection. The combination of this general concept with the extensive scripting functionality of the Cactvs Chemoinformatics toolkit results in a system with capabilities notably different and more flexible than standard pipelining systems.

Reference

1. Ihlenfeldt, Takahashi, Abe: *Proc. 28th Annual Hawaii International Conference on System Sciences* 1995, 227-236.

P16

Asymmetric transfer hydrogenation of imines and ketones using chiral Ru(II)Cl(η^6-*p*-cymene)[(*S,S*)-*N*-TsDPEN] catalyst: a computational study

Petr Kačer[1*], Jiří Václavík[1], Jan Přech[1], Marek Kuzma[2]
[1]ICT – Prague, Technická 5, 166 28 Prague, Czech Republic; [2]Institute of Microbiology, Academy of Sciences, Vídeňská 1083, 142 20, Prague 4, Czech Republic
E-mail: Petr.Kacer@vscht.cz
Journal of Cheminformatics 2012, **4(Suppl 1):**P16

Noyori et al. in 1996 showed that the Ru(II)Cl(η^6-*p*-cymene)[*N*-*p*-tosyl-1,2-diphenylethylenediamine] (= [RuCl(η^6-*p*-cymene)TsDPEN]) in a HCOOH / triethylamine (TEA) mixture was able to efficiently hydrogenate substituted isoquinolines with high enantioselectivity (asymmetric transfer hydrogenation (ATH)) [1]. Almost simultaneously, the same system was reported to reduce ketones superbly by *Fujii et al.* (1996), suggesting that the mechanisms of ATH of C=N and C=O bonds should be alike [2]. The mechanism of the asymmetric reduction of ketones was extensively discussed by *Noyori* in a computational study in 2001, which proposed that the reaction proceeds *via* six-membered transition states (TSs) in the outer coordination sphere of ruthenium [3]. However, this mechanistic concept is not compatible with the ATH of imines, as pointed out by *Martins et al.* in 2009 [4]. According to the original mechanism, the (*S,S*)-complex would give an (*S*)-configured product, which conforms with the results for ketones but disagrees with experimental observations for the ATH of imines. The key element explaining this contrast seems to be the fact that an imine can only be reduced under acidic conditions, which supports the notion of requisite imine protonation, even though this is still not entirely confirmed. In the present work density functional theory (DFT) computational methods were used to investigate the increasingly popular ionic mechanistic concept for the ATH of imines on the Ru(II)Cl (η^6-*p*-cymene)[(*S,S*)-*N*-*p*-tosyl-1,2-diphenylethylenediamine] chiral catalyst. Applying the ionic mechanism, the reaction preferentially affords the (*R*)-amine product, which is in agreement with the experimental observations. Calculated transition state structures for the hydrogenation of protonated 1-methyl-3,4-dihydroisoquinoline are discussed together with their preceding and following energy minima. Stabilization of the favorable transition state by a CH/π interaction between the η^6-*p*-cymene ligand and the substrate molecule is explored in depth to show that both C(sp^2)H/π is more probable than C(sp^3)H/π in this molecular system. Finally, transition state geometries for the ATH of acetophenone are proposed, which take the "standard" six-membered cyclic form.

Acknowledgement: This work has been financially supported by the Grant Agency of the Czech Republic (Grant GACR 104/09/1497), the Ministry of Education, Youth, and Sports of the Czech Republic (Grant CEZ: MSM 604 613 7301). The access to the MetaCentrum computing facilities, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" LM2010005 funded by MEYS.

References

1. Noyori R, *et al*: *Acc Chem Res* 1997, **30**:97.
2. Fujii A, *et al*: *J Am Chem Soc* 1996, **11**:2521.
3. Noyori R, *et al*: *Org Chem* 2001, **66**:7931.
4. Martins JED, *et al*: *Org Lett* 2009, **11**:847.

P17

http://Mcule.com: a public web service for drug discovery

Robert Kiss*, Mark Sandor, Ferenc A Szalai
mcule.com Ltd., Budapest, H-1096, Hungary
E-mail: rkiss@mcule.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P17

http://Mcule.com - a web service providing you a fast and cost-effective way to identify and order new drug candidates has been recently launched. The service is available for the public and it provides a comprehensive, carefully curated database of molecules immediately available for virtual screening. Several screening tools have been already implemented and more will be added on a weekly/monthly basis. Screening tools can be seamlessly integrated into a virtual screening workflow. Calculations are running on cloud machines providing a practically infinite number of CPUs and thus fast access to the screening results. Hits from the virtual screens can be ordered.

P18

COSMOsim3D for drug-similarity, alignment, and molecular field analysis

Andreas Klamt[1*], Karin Wichmann[1], Michael Thormann[2]
[1]COSMOlogic GmbH&CoKG, Leverkusen, and Inst. of Physical and Theoretical Chemistry, University of Regensburg, Germany; [2]Origenis GmbH, Martinsried, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P18

By the practical success of the COSMO-RS fluid phase thermodynamics model [1,2] in many areas of chemistry, the COSMO polarization charge densities σ have been proven to be excellent descriptors for the quantification of the most important kinds of molecular interactions in the liquid phase, as polar interactions, hydrogen bonding and hydrophobicity. In several comparisons and blind tests COSMO-RS predictions for free energies and enthalpies of molecules in solution have been shown to be more accurate than those of most other methods in used in computational chemistry.

Since the same intermolecular interaction modes, which govern fluid phase thermodynamics, are also responsible for binding of ligands to receptors, it is most plausible that a σ-based description of of ligand-ligand similarity or of ligand-receptor interactions should be very promising. Yet, since in COSMO-RS theory the 3D- distribution of the polarization charge density σ on the molecular surface is reduced to a histogram, the σ-profile, the initial approaches in this direction [3] mostly have disregarded the spatial distribution of surface polarities. Only recently we have achieved a 3D-representation of the surface polarization charges by forming generating local σ-profiles on a regular grid. Based on their local σ-profiles, a 3D σ-similarity of two molecules can be defined as the sum of the σ-similarities on the grid points and by optimizing this similarity through rotation and translation of the probe molecule versus the target. This method is introduced here as COSMOsim3D. The quality of this COSMOsim3D similarity measure is demonstrated by a large scale evaluation on bio-isosters.

The same technique can be used to consistently align a set of ligand molecules based on their local σ-profiles. After such alignment, the local σ-profiles or the set of local σ-moments, which is a compressed representation of the σ-profiles, are available as a grid-based local descriptor set, which can be used for subsequent molecular field analysis (MFA). COSMO-RS theory even provides a rationale that the binding free energy, including desolvation, must be a linear function of this set of descriptors. First validation studies have proven that the COSMOsim3D based MFA yields very promising results.

References

1. Klamt A: *J Phys Chem* 1995, **99**:2224-2235.
2. Klamt A: **COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamic and Drug Design.** Amsterdam: Elsevier 2005.
3. Thormann M, Klamt A, Hornig M, Almstetter M: **COSMOsim: Bioisosteric Similarity Based on COSMO-RS σ Profiles.** *J Chem Inf Model* 2006, **46**:1040-1053.

P19

Large scale chemical patent mining with UIMA and UNICORE

Alexander Klenner[1*], Sandra Bergmann[2], Marc Zimmermann[1], Mathilde Romberg[2]
[1]Fraunhofer-Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, 53754, Germany; [2]Forschungszentrum Juelich GmbH, Juelich, 52425, Germany
E-mail: klenner@scai.fraunhofer.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P19

Finding information about annotated chemical reactions for drugs and small compounds is a crucial step for pharmaceutical industries. This data often is presented in form of unstructured documents (especially patents) and manual extraction of this information is a time- and cost inefficient effort.

In our project UIMA-HPC [1], we describe the combined usage of Unstructured Information Managment Architecture (UIMA) and Uniform

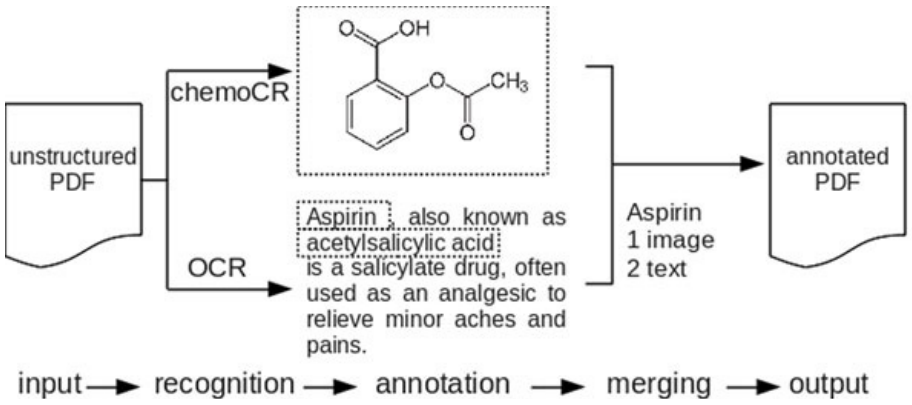

Figure 1(abstract P19) Planned workflow of our UIMA framework. 'Recognition' and 'annotation' are CPU intensive parts that are parallelized on demand using the UNICORE framework. 'Merging' checks for cross-annotations (entity in text and image). Finally, an annotated PDF is presented as output.

Interface to Computing Recources (UNICORE) for large-scale chemical patent mining. Our approach will incorporate existing software such as chemoCR for image processing (image-to-structure) and OCR for text reconstruction. All components are wrapped inside the UIMA framework pipeline. Using the UIMA framework ensures compatibility between different components of the pipeline and makes it possible to connect arbitrary annotation modules into this system. Scale-out for large document collections is achieved by the UNICORE framework on High Performance Clusters, which enables parallelization of all UIMA nodes. The aim is a fully annotated pdf collection where all biomedical entities (compound names, reaction schemes, etc.) are connected by references and thus can be easily browsed and searched by the user. Planned schematic workflow is shown in Figure 1.

Funding: BMBF grant 01IH1101.

Reference

1. [http://www.uima-hpc.org].

P20

CavKA^{HYBRID} – between hard spheres and Gaussians

Florian Koelling*, Knut Baumann
Institut für Medizinische und Pharmazeutische Chemie, Technische Universität Braunschweig, Beethovenstr. 55, D-38106 Braunschweig, Germany
E-mail: florian.koelling@tu-bs.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P20

Three dimensional pharmacophores and pharmacophore searches are well established in virtual screening and have been applied successfully in many prospective and retrospective drug discovery campaigns [1]. While the pharmacophore concept offers an easy and abstract understanding of molecular properties, plenty of user intervention is required to define feasible models.

Recently, Silicos NV provided their freely available ligand centric pharmacophore method Pharao [2]. Pharao employs three dimensional Gaussians to reflect a molecule's pharmacophoric properties, in contrast to most methods which use conventional hard sphere models. Gaussian models show the advantage that they require far less user intervention for model creation.

Here we present a hybrid model that utilizes the Gaussian pharmacophore representation of Pharao to be adapted and used in CavKA (Cavity Knowledge Acceleration), our own in-house strategy for structure based pharmacophore generation. CavKA interprets ligand-receptor complexes and detects interaction between ligand and binding site in order derive pharmacophore models automatically. In addition Grid [3] Molecular Interaction Fields (MIFs) can be used to weight and prioritize interacting features. By combining the smooth nature of Gaussian pharmacophores in the binding site and representing the receptor by a hard sphere excluded volume, structure based pharmacophores can be created without any user intervention.

The performance of CavKA^{HYBRID} is compared to the purely ligand centric approach Pharao in a retrospective virtual screening on the Field Screen [4] dataset. It is the objective of this study to analyse in how far protein information can be used to improve screening results while no user intervention is required for model building. Furthermore the effect of a Gaussian representation of the excluded volume in CavKA^{HYBRID} is analysed.

References

1. Leach AR, Gillet VJ, Lewis RA, Tailor R: **Three-Dimensional Pharmacophore Methods in Drug Discovery.** *J Med Chem* 2010, **53**:539-558.
2. **Pharao 3.0.3, Silicos NV.** [http://www.silicos.be].
3. **Grid 22a, Molecular Discovery.** [http://www.moldiscovery.com].
4. Cheeseright TJ, Mackey MD, Melville JM, Vinter JG: **FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set.** *J Chem Inf Model* 2008, **48**:2108-2117.

P21

MoSGrid: efficient data management and a standardized data exchange format for molecular simulations in a grid environment

Georg Birkenheuer[1], Dirk Blunk[2], Sebastian Breuers[2], André Brinkmann[1], Ines dos Santos Vieira[3], Gregor Fels[1], Sandra Gesing[4], Richard Grunzke[5], Sonja Herres-Pawlis[3], Oliver Kohlbacher[4*], Jens Krüger[1], Ulrich Lang[2], Lars Packschies[2], Ralph Müller-Pfefferkorn[5], Patrick Schäfer[6], Thomas Steinke[6], Klaus-Dieter Warzecha[2], Martin Wewior[2]
[1]Universität Paderborn, Paderborn 33098, Germany; [2]Universität zu Köln, Köln 50923, Germany; [3]Fakultät Chemie, Technische Universität Dortmund, Dortmund 44221, Germany; [4]Eberhard-Karls-Universität Tübingen, Tübingen 72074, Germany; [5]Technische Universität Dresden, Dresden 01187, Germany; [6]Konrad-Zuse-Zentrum für Informationstechnik, Berlin 14195, Germany
E-mail: oliver.kohlbacher@uni-tuebingen.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P21

The MoSGrid (Molecular Simulation Grid) project is currently establishing a platform that aims to be used by both experienced and inexperienced researchers to submit molecular simulation calculations, monitor their progress, and retrieve the results. It provides a web-based portal to easily set up, run, and evaluate molecular simulations carried out on D-Grid resources. The range of applications available encompasses quantum chemistry, molecular dynamics, and protein-ligand docking codes.

In addition, data repositories were developed, which contain the results of calculations as well as "recipes" or workflows. These can be used, improved, and distributed by the users. A distributed high-throughput file system allows efficient access to large amounts of data in the repositories. For storing both the input and output of the calculations, we have developed

MSML (Molecular Simulation Markup Language), a CML derivative (Chemical Markup Language). MSML has been designed to store structural information on small as well as large molecules and results from various molecular simulation tools and docking tools. It ensures interoperability of different tools through a consistent data representation.
At http://www.mosgrid.de the new platform is already available to the scientific community in a beta test phase. Currently, portlets for generic workflows, Gaussian, and Gromacs applications are publicly accessible [1,2].

References

1. Gesing, *et al*: A Science Gateway for Molecular Simulations. *EGI User Forum 2011* 978 90 816927 1 7 2011, 94-95, Book of Abstracts.
2. Gesing, *et al*: Granular Security for a Science Gateway in Structural Bioinformatics. *IWSG-Life 2011 (International Workshop on Science Gateways for Life Sciences)* London, UK 2011.

P22

From eScience to iScience "I want Answers not Links" – new ways to search the Internet

AJ Kos*, D Steinen
Dr. Alexander Kos, AKos GmbH, Austr. 26, D-79585 Steinen, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P22

The industry uses the word eScience to illustrate that in most parts of science everything is done with computers. In chemistry, computer simulations and all the multitude of software applications has reduced the need for many experiments. This is good in a world of diminishing resources. But one application is slow to get a foothold in academia, it is the electronic lab journal. This is very unfortunate because this application is probably the one that makes writing publications and Ph.D. theses much quicker. This is a personal benefit, and a global benefit is that it makes accessible all these experiments that did not work, or did not make it into publication. This information is by far the largest amount of knowledge, and this is wasted, today. This an area where we need to catch up.
Another area where chemists need to rethink is how to access the wealth of information in the Internet. We chemist are all using Google or similar search engines, but we cannot search in the language of the chemists, the chemical structure. CWM Global Search solves this problem. It allows searching Google and more than 50 other major Internet sources by structure, CAS Registry Number, name and identifier. This is very good, but what we get are still links. From there we have to click many times until we finally find an answer. Do you sometimes limit your search to Wikipedia, because you can quickly and directly find an answer? But, Wikipedia is not the Internet. For scientist the answer is often not a single fact, but a table, or even better a graph that visualizes the data. It seems this vision is far in the future.
We cannot solve everything, but with Pipeline Pilot and similar workflow programs, we can build solutions that extract data from the link pages.
The Documents and Text Collection (DTC) for Pipeline Pilot focuses on finding, analyzing, and displaying information from text and documents. A key use of the DTC is to enhance and support the analysis of experimental data generated during your research projects. To do this, the DTC offers broad capabilities, including the ability to search and retrieve documents from online (e.g., PubMed, US & WO Patents) and local (e.g., SharePoint) repositories, and to crawl web sites to gather supporting data and competitive intelligence. For deeper text analytic methods DTC also integrate with 3rd party text analytics applications and technologies such as Linguamatics I2E and the UIMA framework.

P23

CLOUD – CeMM library of unique drugs

Patrick Markt*, Gerhard Dürnberger, Jacques Colinge, Stefan Kubicek
CeMM Research Center for Molecular Medicine, Lazarettgasse 14/AKH BT 25.3, 1090 Vienna, Austria
Journal of Cheminformatics 2012, **4(Suppl 1):**P23

Approved drugs serve as functional probes for analyzing disease-relevant biochemical pathways or are screened against proteins to identify novel medical applications resulting in a repurposing of known drugs. Despite these benefits of approved drugs, it is almost impossible to physically obtain a complete collection. Commercial vendors cover about 50-65% of drugs. Academic efforts such as the Johns Hopkins Clinical Compound Library (JHCCL) [1] and the NCGC Pharmaceutical collection [2] are only accessible via collaboration agreements. Thus, an affordable reference drug library obtainable for all screening centers is still missing. Here we report the generation of the CeMM Library of Unique Drugs (CLOUD), a focused collection of 314 systemically bioavailable small molecule drugs. The library was derived from FDA-approved drugs applying data mining and structural clustering techniques. Reduction of approved drugs to a set of systemically available small molecules and clustering based on their activity classes (e.g. dopamine receptor agonist) resulted in 244 compounds. In addition to this representative set of structural and biological space of drugs, CLOUD contains 35 drugs where the biological target is still unknown. Another 35 molecules representing the active forms of CLOUD prodrugs ensure the usability of the library for both biochemical and cell-based assays. Finally, all CLOUD drugs are delivered in a single 384-well plate in concentrations related to their therapeutic plasma levels to make high-throughput screening as comfortable as possible.

References

1. Chong CR, Chen X, Shi L, Liu JO, Sullivan DJ Jr: A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat Chem Biol* 2006, **2**:415-416.
2. Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, Nguyen DT, Austin CP: The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. *Sci Transl Med* 2011, **3**:80ps16.

P24

Activity-difference maps and consensus similarity measure characterize structure-activity relationships

José L Medina-Franco[1*], Austin B Yongye[1], Jaime Pérez-Villanueva[2], Richard A Houghten[1], Karina Martínez-Mayorga[1]
[1]Torrey Pines Institute for Molecular Studies, Port St. Lucie, Florida, 34987, USA; [2]Departament of Biological Systems, UAM-X, Mexico City, 04960, Mexico
E-mail: jmedina@tpims.org
Journal of Cheminformatics 2012, **4(Suppl 1):**P24

Dual and triple activity-difference (DAD/TAD) maps are two- and three-dimensional representations of the pairwise activity differences of compound data sets, respectively [1]. These maps are valuable tools for the systematic characterization of structure-activity relationships (SAR) of compounds data sets screened against two or three targets [2]. Adding pairwise structural similarity information into the DAD/TAD maps readily reveals *activity cliff* [3] regions in the SAR for one, two or the three targets. In addition, pairs of compounds in the smooth regions of the SAR and scaffold hops are also easily identified in these maps. Herein, we describe DAD and TAD maps for the systematic characterization of the SAR of data sets screened against three molecular targets. Several 2D and 3D structure representations were used to characterize the SAR in order to reduce the well-known dependence of the activity landscape on the structural representation [4,5]. Systematic analysis of the DAD and TAD maps reveals regions in the landscape with similar SAR for two or the tree targets as well as regions with inverse SAR, i.e., changes in structure that increase activity for one target, but decrease activity for the other target. Focusing the analysis on pairs of compounds with high structure similarity revealed the presence of single-, dual- and triple-target activity cliffs, i.e., small changes in structure with high changes in potency for one, two or the three targets, respectively. Triple-target scaffold hops are also discussed.

References

1. Pérez-Villanueva J, Santos R, Hernández-Campos A, Giulianotti MA, Castillo R, Medina-Franco JL: Structure-Activity Relationships of Benzimidazole Derivatives as Antiparasitic Agents: Dual Activity-Difference (DAD) Maps. *Med Chem Comm* 2011, **2**:44-49.
2. Dimova D, Wawer M, Wassermann AM, Bajorath J: Design of Multitarget Activity Landscapes that Capture Hierarchical Activity Cliff Distributions. *J Chem Inf Model* 2011, **51**:258-266.

3. Maggiora GM: **On Outliers and Activity Cliffs-Why QSAR Often Disappoints.** *J Chem Inf Model* 2006, **46**:1535-1535.
4. Medina-Franco JL, Martínez-Mayorga K, Bender A, Marín RM, Giulianotti MA, Pinilla C, Houghten RA: **Characterization of Activity Landscapes Using 2D and 3D Similarity Methods: Consensus Activity Cliffs.** *J Chem Inf Model* 2009, **49**:477-491.
5. Yongye A, Byler K, Santos R, Martínez-Mayorga K, Maggiora GM, Medina-Franco JL: **Consensus Models of Activity Landscapes with Multiple Chemical, Conformer and Property Representations.** *J Chem Inf Model* 2011, **51**:1259-1270.

P25

Computer-guided discovery of epigenetics drugs: molecular modeling and identification of inhibitors of DNMT1

Jakyung Yoo*, José L Medina-Franco
Torrey Pines Institute for Molecular Studies, Port St. Lucie, Florida, 34987, USA
Journal of Cheminformatics 2012, **4(Suppl 1):**P25

DNA methylation is a covalent chemical modification of DNA catalyzed by DNA methyltransferases (DNMTs) and plays a crucial role in epigenetic modifications. Inhibition of DNMT is a promising strategy for the treatment of various developmental and proliferative diseases, particularly cancers. Molecular docking and other computational approaches are increasingly being used to explore the ligand-binding interactions of DNMT inhibitors [1,2]. In this work we conducted molecular docking of experimentally known active compounds in the catalytic site of the recently published crystal structure of DNMT1 [3]. Prior docking, the conformation of the catalytic site was modelled with molecular dynamics into an active conformation. To our knowledge, this is the first molecular modelling study conducted with the catalytic binding site of this crystal structure. Based on the docking results, we developed a structure-based pharmacophore model. Molecular modelling results were compared with the insights previously obtained with a homology model of the methyltransferase domain of DNMT1 [4]. We also discuss the experimental inhibitory activity and docking of a novel DNMT1 inhibitor recently identified in our group [5]. Results of this work have direct implications in the future computer-based screening and optimization of inhibitors of DNMT1 and show that computational approaches form part of multidisciplinary efforts to further advance epigenetic therapies [1].

References

1. Medina-Franco JL, Caulfield T: **Advances in the Computational Development of DNA Methyltransferase Inhibitors.** *Drug Discovery Today* 2011, **16**:418-425.
2. Kuck D, Caulfield T, Lyko F, Medina-Franco JL: **Nanaomycin A Selectively Inhibits DNMT3B and Reactivates Silenced Tumor Suppressor Genes in Human Cancer Cells.** *Mol Cancer Ther* 2010, **9**:3015-3023.
3. Song J, Rechkoblit O, Bestor TH, Patel DJ: **Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation.** *Science* 2011, **331**:1036-1040.
4. Yoo J, Medina-Franco JL: **Homology Modeling, Docking, and Structure-based Pharmacophore of Inhibitors of DNA Methyltransferase.** *J Comp-Aided Mol Des* 2011, **25**:555-567.
5. Yoo J, Medina-Franco JL: **Trimethylaurintricarboxylic Acid Inhibits Human DNA Methyltransferase 1: Insights from Enzymatic and Molecular Modeling Studies.** *J Mol Model* in press, doi:10.1007/s00894-011-1191-4.

P26

A large scale classification of molecular fingerprints for the chemical space representation and SAR analysis

Fabian López-Vallejo*, Jacob Waddell, Austin B Yongye, Richard A Houghten, José L Medina-Franco
Torrey Pines Institute for Molecular Studies, Port St. Lucie, Florida, 34987, USA
Journal of Cheminformatics 2012, **4(Suppl 1):**P26

Fingerprint-based structure representation has a broad range of applications including, but not limited to, diversity analysis, compound classification, chemical space visualization [1], activity landscape modelling and similarity searching. It has been shown that depending on the particular fingerprints used, the outcome of similarity searching [2] or activity landscapes [3] can be very different. Combining structure representations is a common practice to increase the performance of similarity searching [4]. Also, combining representations for activity landscape modelling has been proposed to generate robust descriptive SAR models [5]. However, the selection of fingerprints to be combined is not an easy task. As part of our efforts to select fingerprint representations to generate consensus representations of chemical space and activity landscapes [5,6] herein we discuss the results of a systematic comparison of more than 10 2D and 3D fingerprint representations in terms of performance in diversity analysis (as opposed to similarity searching). We employed more than 20 data sets from different sources relevant to drug discovery. In this work the widely used Tanimoto coefficient was employed. The approach presented here can be easily extended to other similarity measures, additional fingerprints and molecular databases. We also discuss the typical mean/median similarity values of selected fingerprints across databases from different sources.

References

1. Medina-Franco JL, Martínez-Mayorga K, Giulianotti MA, Houghten RA, Pinilla C: **Visualization of the Chemical Space in Drug Discovery.** *Curr Comput-Aided Drug Des* 2008, **4**:322-333.
2. Bender A: **How Similar Are Those Molecules after All? Use Two Descriptors and You Will Have Three Different Answers.** *Expert Opin Drug Discovery* 2010, **5**:1141-1151.
3. Wassermann AM, Wawer M, Bajorath J: **Activity Landscape Representations for Structure-Activity Relationship Analysis.** *J Med Chem* 2010, **53**:8209-8223.
4. Chen B, Mueller C, Willett P: **Combination Rules for Group Fusion in Similarity-Based Virtual Screening.** *Mol Inf* 2010, **29**:533-541.
5. Yongye A, Byler K, Santos R, Martínez-Mayorga K, Maggiora GM, Medina-Franco JL: **Consensus Models of Activity Landscapes with Multiple Chemical, Conformer and Property Representations.** *J Chem Inf Model* 2011, **51**:1259-1270.
6. López-Vallejo F, Nefzi A, Bender A, Owen JR, Nabney IT, Houghten RA, Medina-Franco JL: **Increased Diversity of Libraries from Libraries: Chemoinformatic Analysis of Bis-Diazacyclic Libraries.** *Chem Biol Drug Des* 2011, **77**:328-342.

P27

Computer-aided studies of molecular structure-comparison of measured and computed ECD spectra

Richard Moha*, Verena Gossen
Institute of Organic Chemistry, RWTH Aachen University, Landoltweg 1, 52074 Aachen, Germany
E-mail: Richard.Moha@rwth-aachen.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P27

ECD spectroscopy is an important instrument not only for the analysis of chirality but also for the study of conformational aspects of organic compounds [1]. In the present case, we studied helical complexes with different metals in the Λ or Δ form. ECD spectra of the complexes measured in DMSO (Figure 1) are compared to their calculated counterparts revealing conformational aspects of the structure. The structures of the complexes were optimized with the CAM-B3LYP functional [2], the SDD basis set and the effective core potential for the metal as implemented in the program package Gaussian09 [3]. TZVP was used for all other atoms. Subsequent time dependent DFT calculations were performed with the B3LYP functional.

Depending on the solvent and concentration, measured CD spectra are more or less noisy. Moreover, calculated spectra are frequently shifted to the blue while measured and calculated Δε might differ which complicates the analyses. Our newly coded *Spectra Curve Manager* [4] partly solves these problems. First the program subtracts the background from the measured raw spectrum. Subsequently, a graphical interface analyses differences between the measured and calculated spectra. Smoothing of the experimental spectra is performed by numerical algorithms, and the shift of the spectrum is minimized by displacement of the calculated spectra along the λ-axis. The Δε can be fitted as well. These combined methods facilitate comparison of measured and calculated spectra and, therefore, analysis of experimental results.

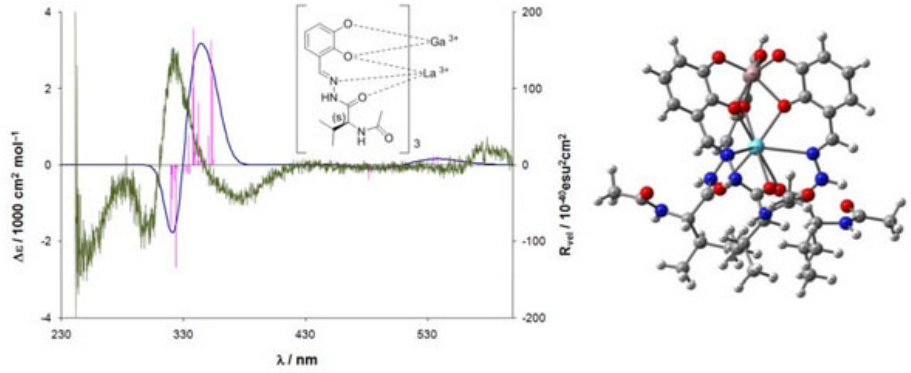

Figure 1(abstract P27) Left: Comparison of the measured CD spectrum in DMSO and the calculated spectrum of the complex, right: geometrically optimized structure (CAM-B3LYP).

References

1. Albrecht M, Isaak E, Baumert M, Gossen V, Raabe G, Fröhlich R: **.** *Angew Chem Int Ed* 2011, **50**:2850.
2. Yanai T, Tew D, Handy N: **.** *Chem Phys Lett* 2004, **393**:51.
3. **Gaussian 09, Revision A02.** Gaussian Inc., Wallingford CT 2009.
4. Moha R: **Spectra Curve Manager.** RWTH Aachen 2011.

P28

Novel binding pocket descriptors based on DrugScore potential fields encoded by 3D Zernike descriptors

Britta Nisius*, Fan Sha, Holger Gohlke
Department of Mathematics and Natural Sciences, Institute of Pharmaceutical and Medicinal Chemistry, Heinrich-Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P28

Proteins interact with other molecules, e.g. ligands or other proteins, in specific binding sites. Key factors for these interactions are the shape, size, and buriedness of the binding site, as well as its physicochemical composition. Since all these properties usually significantly vary among different proteins, up to now there is no standard definition what constitutes a binding site [1]. Thus, novel pocket descriptors allowing an in-depth characterization of binding sites are highly desired.
Hence, we developed novel binding pocket descriptors based on 3D molecular interaction fields. The binding pocket of a protein is characterized using the distance dependent, knowledge-based pair potentials of the DrugScore scoring function [2] in combination with multiple ligand atom probes. To allow an efficient comparison of the resulting potential fields, the 3D grids are encoded using 3D Zernike descriptors.
The 3D Zernike polynomials Z_{nlm} are orthonormal basis functions on the unit sphere. Thus, any 3D object can be represented as:

$$f(x,y,z) = \sum_{n=0}^{\infty} \sum_{l=0}^{n} \sum_{m=-l}^{l} c_{nlm} \cdot Z_{nlm}(x,y,z)$$

using a 3D Zernike function expansion. We utilized the resulting function expansion coefficients c_{nlm}, i.e. the 3D Zernike moments, to describe the 3D molecular potential fields characterizing a protein's binding pocket.
The resulting descriptors are invariant under rotation, scaling, and translation and enable a fast comparison and an efficient characterization of protein binding pockets. Thus, these novel pocket descriptors can be used to predict target druggability or to calculate similarities between binding pockets, e.g. to predict potential off-targets or to perform protein function de-orphanization.

References

1. Perot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO: **Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery.** *Drug Discov Today* 2010, **15**:656-667.
2. Gohlke H, Hendlich M, Klebe G: **Knowledge-based Scoring Function to Predict Protein-Ligand Interactions.** *J Mol Biol* 2000, **295**:337-356.

P29

Computational studies of flaviviruses: approaching to novel fusion inhibitors

Dmitry I Osolodkin[1,2,3*], Liubov I Kozlovskaya[2], Galina G Karganova[2], Evgenia V Dueva[1], Vladimir A Palyulin[1,3], Nikolay S Zefirov[1,3], Vladimir M Pentkovski[3]
[1]Department of Chemistry, Moscow State University, Moscow 119991, Russia; [2]Chumakov Institute of Poliomyelitis and Viral Encephalitides RAMS, Moscow region 142782, Russia; [3]Moscow Institute of Physics and Technology, Dolgoprundny 141700, Russia
E-mail: dmitry_o@qsar.chem.msu.ru
Journal of Cheminformatics 2012, **4(Suppl 1):**P29

Flaviviral infections affect millions of people throughout the world, in such different territories as Europe, Australia, USA, and China. Among these infections are dengue fever, tick-borne encephalitis, Powassan encephalitis, West Nile fever, yellow fever, Omsk haemorrhagic fever, etc. Effective vaccines exist against only several flaviviruses. Moreover, there is no other specific therapy for flaviviral infections; consequently, new antiviral drugs are needed for the treatment of these diseases.
Falviviruses are characterised by enveloped virion with E protein forming its outer surface. During the entry to the target cell E protein undergoes pH-induced conformational switch into fusogenic state that drives fusion of virion and cell membranes with consequent release of the viral genome into cytoplasm.
Our studies are devoted to the modelling of flavivirus virion envelope proteins structure and molecular dynamics simulation, generally aiming in design of the fusion inhibitors. First, we studied ectodomain of TBEV E protein and revealed the possible mechanism of differences in virus binding to glycosaminoglycans [1]. Second, the analysis of point substitutions in TBEV E protein ectodomain in MD simulation study revealed the basis of their influence on the virion properties. These studies were expanded to include the stem and anchor region of this protein and describe its interactions with the viral membrane. The global aim of this study is an atom-scale simulation of viral fusion process.
Another aspect of the fusion inhibitor design is analysis of the inhibitor binding site composition and structure-based virtual screening of the putative inhibitors. Such analysis has been performed for DENV, TBEV and POWV and allowed identifying molecules with high probability to inhibit the crucial step of these infections.

Reference

1. Kozlovskaya LI, Osolodkin DI, Shevtsova AS, Romanova LI, Rogova YV, Dzhivanian TI, Lyapustin VN, Pivanova GP, Gmyl AP, Palyulin VA, Karganova GG: **GAG-binding variants of tick-borne encephalitis virus.** *Virology* 2010, **398**:262-272.

P30

Identification of hot-spot regions of N-type Ca^{2+} channel receptor by homology modeling and molecular dynamics study, for structure-based blocker design

Ashish Pandey*, Satyaprakash Tripathi, C Gopi Mohan
Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER),Sector 67, S.A.S. Nagar,Punjab -160 062, India
E-mail: ashishpandey12@gmail.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P30

The voltage dependent N-type Ca^{2+} channel (NCC) distributed in the nerve ending of the central and peripheral nerves. NCC is considered as potential therapeutic target for several pathological disorders like neuropathic pain and stroke disease. For understanding mechanism of action at the atomic level crystal structure provide valuable inside but lack of crystal structure of ion channel lead sequence analysis of different types of voltage dependent Ca^{2+} channel (VDCC) and we found identical/similar active site residues, which was confirmed by site-directed mutagenesis analysis of L-type Ca^{2+} channel (LCC). Based on these observations, we have developed for the first time the homology model of the closed state of NCC receptor at the ligand-sensing region by using bacterial K^+ channel receptor. Further, molecular docking using different dihydropyridine (DHP) blockers identified NCC receptor hot spot binding residues, which is in consonance with that of the LCC. These residues are potential for further biochemical investigations. To understand binding and stability behavior of NCC with the DHP (amlodipine) in closed state, 50 nano second molecular dynamics simulation in lipid bilayer membrane environment were carried out. This analysis revealed the closed state stabilizing by binding of ligand into inner part of S6 region.

P31

SAR Analyzer: a tool for interactive SAR data visualization and navigation

Lisa Peltason*, Daniel Stoffler
Therapeutic Modalities Informatics / Cheminformatics and Statistics, Pharma Research & Early Development Informatics, F. Hoffmann-La Roche Ltd, CH-4070 Basel, Switzerland
E-mail: lisa.peltason@roche.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P31

The analysis of structure-activity relationships (SAR) is a central task in medicinal chemistry. The association between chemical structures of biologically active molecules and multiple property and assay data provides the basis for selection, optimization, and evaluation of potential drug candidate molecules.

In an endeavour to facilitate the navigation of complex data landscapes and enable intuitive access to SAR information, an interactive SAR analysis platform is being developed at Roche. Focusing on information-rich data visualizations, we aim at supporting the medicinal chemists in their decision-making process, rather than making quantitative or qualitative predictions. The "SAR Analyzer" integrates cutting-edge visualization techniques and enhances them with key capabilities such as real-time interaction with the data, an intuitive user interface, and functionality to easily extract and navigate multi-property data.

Here we show how two complementary approaches are integrated in the SAR Analyzer. The "SAR Map" visualization provides a holistic view of the distribution of molecular structures and properties in a data set [1]. Molecules are displayed in a 2D map projection based on chemical features or similarity. Color shading indicates the distribution of selected molecular properties or biological activity. By contrast, the "SAR Tree" visualization [2] focuses on subsets of similar compounds. Starting at a user-selected reference compound, all molecules within a similarity radius are organized in a hierarchical tree structure that makes it possible to interactively browse different compound series and derive and test SAR hypotheses.

Both concepts depart from classical SAR data analysis and support scientific reasoning by making SAR information accessible in an intuitive visual way. Linking both approaches allows the analysis of SAR data on different levels of detail and helps to address questions that are relevant in different stages of a medicinal chemistry research project.

References

1. Peltason L, Iyer P, Bajorath J: **Rationalizing three-dimensional activity landscapes and the influence of molecular representations on landscape topology and formation of activity cliffs.** *J Chem Inf Model* 2010, **50**:1021-1033.
2. Wawer M, Bajorath J: **Similarity-potency trees: a method to search for SAR information in compound data sets and derive SAR rules.** *J Chem Inf Model* 2010, **50**:1395-1409.

P32

Molecular dynamics simulations and docking of non-nucleoside reverse transcriptase inhibitors (NNRTIs): a possible approach to personalized HIV treatment

Florian D Roessler[1,3*], Oliver Korb[2], Andreas Bender[1], Werner Maentele[3], Peter J Bond[1]
[1]Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge,CB2 1EW, UK; [2]The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK; [3]Institute of Biophysics, Johann Wolfgang Goethe-University Frankfurt, Max von Laue-Strasse 1, 60438 Frankfurt am Main, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P32

The human immunodeficiency virus (HIV) is currently ranked sixth in the worldwide causes of death [1]. One treatment approach is to inhibit reverse transcriptase (RT), an enzyme essential for reverse transcription of viral RNA into DNA before integration into the host genome [2]. By using non-nucleoside RT inhibitors (NNRTIs) [3], which target an allosteric binding site, major side effects can be evaded. Unfortunately, high genetic variability of HIV in combination with selection pressure introduced by drug treatment enables the virus to develop resistance against this drug class by developing point mutations. This situation necessitates treatment with alternative NNRTIs that target the particular RT mutants encountered in a patient.

Previously, proteochemometric approaches have demonstrated some success in predicting binding of particular NNRTIs to individual mutants; however a structurebased approach may help to further improve the predictive success of such models. Hence, our aim is to rationalize the experimental activity of known NNRTIs against a variety of RT mutants by combining molecular modeling, long-timescale atomistic molecular dynamics (MD) simulation sampling and ensemble docking. Initial control experiments on known inhibitor-RT mutant complexes using this protocol were successful, and the predictivity for further complexes is currently being evaluated. In addition to predictive power, MD simulations of multiple RT mutants are providing fundamental insight into the dynamics of the allosteric NNRTI binding site which is useful for the design of future inhibitors. Overall, work of this type is hoped to contribute to the development of predictive efficacy models for individual patients, and hence towards personalized HIV treatment options.

References

1. WHO: **The top 10 causes of death**.
2. Chen LF, Hoy J, Lewin SR: **Ten years of highly active antiretroviral therapy for HIV infection.** *Med J Australia* 2007, **186**:146-151.
3. Turner BG, Summers MF: **Structural biology of HIV.** *J Mol Biol* 1999, **285**:1-32.

P33

Modeling of molecular atomization energies using machine learning

Matthias Rupp[1*], Alexandre Tkatchenko[2], Klaus-Robert Müller[1], O Anatole von Lilienfeld[3]
[1]Machine Learning Group, Technical University, Berlin, 10587, Germany; [2]Fritz-Haber-Institute, Max-Planck Society, Berlin, 14195, Germany; [3]Argonne National Laboratory, Argonne, Illinois, 60439, USA
E-mail: matthias.rupp@tu-berlin.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P33

Atomization energies are an important measure of chemical stability. Machine learning is used to model atomization energies of a diverse set of organic molecules, based on nuclear charges and atomic positions only [1]. Our scheme maps the problem of solving the molecular time-independent Schrödinger equation onto a non-linear statistical regression

problem. Kernel ridge regression [2] models are trained on and compared to reference atomization energies computed using density functional theory (PBE0 [3] approximation to Kohn-Sham level of theory [4,5]). We use a diagonalized matrix representation of molecules based on the internuclear Coulomb repulsion operator in conjunction with a Gaussian kernel. Validation on a set of over 7000 small organic molecules from the GDB database [6] yields mean absolute error of ~10 kcal/mol, while reducing computational effort by several orders of magnitude. Applicability is demonstrated for prediction of binding energy curves using augmentation samples based on physical limits.

References

1. Rupp M, Tkatchenko A, Müller KR, von Lilienfeld OA: **Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning.** *Physical Review Letters* 2012, **108(5)**.
2. Hastie T, Tibshirani R, Friedman J: **The Elements of Statistical Learning: Data Mining, Inference, and Prediction.** New York: Springer 2009.
3. Perdew JP, Ernzerhof M, Burke K: **Rationale for mixing exact exchange with density functional approximations.** *J Phys Chem* 1996, **105(22)**:9982-9985.
4. Hohenberg P, Kohn W: **Inhomogeneous Electron Gas.** *Phys Rev* 1964, **136(3B)**:B864-B871.
5. Kohn W, Sham LJ: **Self-Consistent Equations Including Exchange and Correlation Effects.** *Phys Rev* 1965, **140(4A)**:A1133-A1138.
6. Blum LC, Reymond JL: **970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13.** *J Am Chem Soc* 2009, **131(25)**:8732-8733.

P34

The assessment of computationally derived protein ensembles in protein-ligand docking

Barbara Sander[1*], Oliver Korb[2], Jason Cole[2], Jonathan W Essex[1]
[1]School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK; [2]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK
E-mail: bs3e09@soton.ac.uk
Journal of Cheminformatics 2012, **4(Suppl 1):**P34

The inclusion of receptor flexibility in protein-ligand docking experiments has become a major research interest in drug discovery [1,2]. One of the possible methods applied is the use of multiple discrete protein conformations, so called ensemble docking [3,4]. With computational techniques like Molecular Dynamics (MD) a large number of different conformations can be generated, not all of which can or should be included in the docking or virtual screening process [5]. The question arises if and how suitable protein conformations can be selected systematically *a priori* based on quantifiable conformational features.

For neuraminidase and cyclin-dependent kinase II (CDK2), snapshots of MD simulation trajectories have been clustered based on different structural properties using a variety of clustering methods. To establish a possible correlation between docking performance and target conformational attributes the clustered snapshots have been subjected to extensive self- and cross-docking experiments as well as virtual screening using the GOLD docking programme. It is shown that conformationally similar snapshots do not necessarily result in a similar docking or virtual screening performance. The selection of the particular structural property on which to base the clustering appears to be the essential problem.

References

1. Carlson HA: **Protein flexibility and drug design: how to hit a moving target.** *Curr Opin Chem Biol* 2002, **6**:447-452.
2. Durrant JD, McCammon JA: **Computer-aided drug-discovery techniques that account for receptor flexibility.** *Curr Opin Pharmacol* 2010, **10**:770-774.
3. Barril X, Morley DS: **Unveiling the full potential of flexible receptor docking using multiple crystallographic structures.** *J Med Chem* 2005, **48**:4432-4443.
4. Huang S-Y, Zou H: **Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking.** *Proteins* 2007, **66**:399-421.
5. Lin J-H, Perryman AL, Schames JR, McCammon JA: **Computational drug design accomodating receptor flexibility: the relaxed complex scheme.** *J Am Chem Soc* 2002, **124**:5632-5633.

P35

Development of target-biased scoring functions for protein-ligand docking

Michael Scharfe*, Martin Pippel, Wolfgang Sippl
Institute of Pharmacy, Pharmaceutical Chemistry, University Halle-Wittenberg, Halle, 06120, Germany
E-mail: michael.scharfe@pharmazie.uni-halle.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P35

Accurate scoring of protein-ligand interactions for docking, binding-affinity prediction and virtual screening campaigns is still challenging. Despite great efforts, the performance of existing scoring functions strongly depends on the target structure under investigation. Recent developments in the direction of target-class-specific scoring methods and machine-learning-based procedures reveal significant performance improvement in binding mode and affinity prediction.

However, there is no open-source framework that combines such sophisticated techniques with molecular docking algorithms to make them simply applicable for virtual screening. Therefore, the aim of this work is the implementation of general tools within the open-source framework ParaDockS [1] to obtain target-class-specific scoring functions. These scoring functions are based on knowledge-based atom-pair potentials which can be used in an additive manner (PMF-Score) to score protein-ligand complexes. Because ParaDockS already includes algorithms for protein-ligand docking, every new obtained scoring function can be immediately applied.

Recently it was shown, that atom-pair potentials are also useful for training machine-learning models like support vector machines or random forests [2]. Such methods circumvent a particular functional form for the scoring function and thereby implicitly capture binding contributions that are hard to model explicitly. Our goal is to implement different machine-learning procedures within the ParaDockS framework to provide further possibilities for scoring protein-ligand complexes.

In a first test and validation study, we applied this workflow on kinase data sets, but in principle it is applicable to every target class with enough structural data. In further studies we want to obtain a set of different scoring functions that are biased to a certain target class and can be used for docking and scoring within ParaDockS.

References

1. Meier, Pippel, Brandt, Sippl, Baldauf: **ParaDockS – A framework for Molecular Docking with Population-Based Metaheuristics.** *J Chem Inf Model* 2010, **50**:879-889.
2. Li, Khanna, Jo, Wang, Ashpole, Hudmon, Meroueh: **Target-Specific Support Vector Machine Scoring in Structure-Based Virtual Screening: Computational Validation, In Vitro Testing in Kinases, and Effects on Lung Cancer Cell Proliferation.** *J Chem Inf Model* 2011, **51**:755-759.

P36

Systematic search for pairwise dependencies of torsion angles

Christin Schärfer[1*], Tanja Schulz-Gasch[2], Matthias Rarey[1]
[1]Center for Bioinformatics (ZBH), Bundesstr. 43, 20146 Hamburg, Germany; [2]F. Hoffmann-La Roche Ltd., 4070 Basel, Switzerland
E-mail: schaerfer@zbh.uni-hamburg.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P36

Most available tools for conformer generation, like OMEGA [1], ROTATE [2], and MIMUMBA [3], divide the conformational space into quantized degrees of freedom, i.e. torsion angles, which are treated independently. The independence of torsions is however not valid for all fragments [4]. There are pairs of mutually dependent degrees of freedom e.g. two consecutive torsion angles in aryl-X-aryl systems. The fact that two torsions are dependent implies that if one of the torsions is set to a specific angle, the set of possible angles for the other torsion is limited. This knowledge could be used to significantly narrow down the conformational space in deterministic rule-based conformation generators.

For our systematic search for pairwise dependent torsion angles, we assembled a set of about 200 chemical patterns, each describing a torsion angle and part of its environment. The patterns range from very general descriptions like 'rotatable bond between two sp3 hybridized atoms' to patterns describing a more specific molecular environment.

As a first approach we tried to replicate the examples given by Bramelt et al. using suitable chemical patterns and a CSD [5] subset of about 73,000 molecules as a database. We then performed a pairwise analysis of all our chemical patterns, including each pattern with itself, using again the CSD subset of about 73,000 molecules. We used two different search scenarios. In our first search the torsion angles had to be directly next to each other while in our second search they had to be exactly one bond apart.
Using our systematic search approach we found many additional examples for dependent torsion angles, confirming the findings of Bramelt et al. and supporting their advice to search for pairs of mutually dependent conformation variables.

References

1. OMEGA. OpenEye Scientific Software: Santa Fe, NM, USA 2011.
2. ROTATE. Molecular Networks: Erlangen, Germany 2011.
3. Klebe G, Mietzner M: **A fast and efficient method to generate biologically relevant conformations.** *J Comput-Aided Mol Des* 1994, **8**:583-606.
4. Bramelt KA, Kuhn B, Reuter DC, Stahl M: **Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused Analysis.** *J Chem Inf Model* 2008, **48**:1-24.
5. Allen FH: **The Cambridge Structural Database: a quarter of a million crystal structures and rising.** *Acta Cryst* 2002, **B58**:380-388.

P37

QSPR designer – employ your own descriptors in the automated QSAR modeling process

Ondřej Skřehota[1,2*], Radka Svobodová Vařeková[1,2], Stanislav Geidl[1], Michal Kudera[1], David Sehnal[1], Crina-Maria Ionescu[1], Jan Žídek[3], Jaroslav Koča[1]
[1]NCBR and CEITEC - Central European Institute of Technology, Masaryk University, Brno, CZ 625 00, Czech Republic; [2]BioPeta, Brno, CZ 625 00, Czech Republic; [3]CEITEC, Brno University of Technology, Brno, CZ 616 00, Czech Republic
E-mail: skrehota@chemi.muni.cz
Journal of Cheminformatics 2012, **4(Suppl 1):**P37

The prediction of physical and chemical properties of molecules is a very important step in the drug discovery process. QSAR and QSPR models are strong tools for predicting these properties. The models employ descriptors and statistical approaches to provide an estimation of the desired property. An abundance of descriptors and QSAR/QSPR models were published, but the prediction of some properties (i.e., pK_a, logP) is still a challenge [1]. For this reason, researchers are perpetually working on identifying new descriptors and analyzing their performance in newly designed models. The process of design, parameterization and evaluation of a QSAR/QSPR model is relatively complicated. Therefore, several software tools for its automation are currently under development [2,3]. These tools are very useful if we wish to apply some of many descriptors which they implement. But if we need to use other descriptors or test our own, we require a different solution.
We introduce a new version of the software QSPR Designer that fulfils the above mentioned requirement. Specifically, the user can easily and quickly add his own module for descriptor calculation into QSPR Designer. Alternatively, descriptors can also be obtained from an input file. When the descriptors are available, QSPR Designer allows the user to include them to the model, to validate this model and to perform further tasks. The functionality of QSPR Designer is demonstrated on three case studies, in which the user gradually tests and improves his ideas of how to predict pK_a. The first case study is focused on the question whether pK_a has any relation with atomic charges. The second case study tests if we can use charges and a simple similarity-based model to predict pK_a. The third case study analyzes several QSPR models predicting pK_a from charges.

References

1. Svobodová Vařeková R, Geidl S, Ionescu CM, Skřehota O, Kudera M, Sehnal D, Bouchal T, Abagyan R, Huber JH, Koča J: **Predicting pK_a values of substituted phenols from atomic charges.** *J Chem Inf Model* 2011, **51**:1795-1806.
2. Rodgers LS, Davis AM, Tomkinson NP, van de Waterbeemd H: **Predictivity of Simulated ADME AutoQSAR Models over Time.** *Mol Inf* 2011, **30**:256-266.
3. Luscombe CN, Malcolm N, Cox R, Pickett SD: **QSAR Workbench: Enabling Best Practices in Large Scale Model Building.** *9th ICCS* 2011.

P38

Adsorption of amino acids on MFI-type zeolite: DFT calculations and experimental results

Kai Stueckenschneider[1*], Achim Zielesny[2], Gerhard Schembecker[1]
[1]Department of Biochemical and Chemical Engineering, TU Dortmund, Dortmund, 44227, Germany; [2]Institute for Bioinformatics and Chemoinformatics, University of Applied Sciences Gelsenkirchen, Recklinghausen, 45665, Germany
E-mail: Kai.Stueckenschneider@bci.tu-dortmund.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P38

Adsorption is a common unit operation in separation and purification of biotechnological products where chromatography steps can make up more than half the amount of the total purification costs [1]. The molecular mechanisms of adsorption are still not understood in detail. Further understanding of interactions between adsorbent surfaces and adsorptives could help to facilitate process design in a more cost efficient manner.
In this work the interaction of MFI-type zeolite MFI-27 (Al/Si=13) with Alanine and Phenylalanine is investigated by quantum chemical (QM) calculations which are compared to experimental adsorption data.
For the QM calculations T3-clusters are used as MFI-27 surface models which were shown to be successful e.g. in the case of protolytic cracking of alkanes [2,3]. In order to model different pH-values Alanine and Phenylalanine are applied in their protonated, zwitterionic and deprotonated state. Geometry optimisations and frequency analysis of all molecular structures are performed with Density Functional Theory (DFT) using the B3LYP functional with different basis sets. Calculated complex energies are corrected for BSSE and ZPE. Adsorption isotherms are derived from corresponding experiments.
Regarding the adsorption isotherms it is shown that a high adsorption of Alanine and Phenylalanine on MFI-27 takes place at low pH values (near the pK_a of the amino acids). Less adsorption occurs with an increased pH equalling the amino acids' isoelectric points. At the pH of the amico acids' pK_b values adsorption is no longer observed.
These trends can be correlated with the corresponding QM calculations. High binding energies are calculated for the protonated amino acids. Zwitterionic states lead to lower binding energies. The deprotonated amino acids do not show any binding affinity to MFI-27.
These first results indicate the reliability of the applied methods for our model system.

References

1. Ladisch MR: **Bioseparations Engineering.** Wiley and Sons 2001, 1-5.
2. Zheng X, Blowers P: **Reactivity of Alkanes on Zeolites: A Computational Study of Propane Conversion Reactions.** *J Phys Chem A* 2005, **109**:10734-10741.
3. Zheng X, Blowers P: **Reactivity of Isobutane on Zeolites: A First Principles Study.** *J Phys Chem A* 2006, **110**:2455-2460.

P39

Guiding protein-ligand docking with different experimental NMR-data

Tim ten Brink[1*], Ionut Onila[1], Adam Mazur[2], Oliver Korb[1], Heiko M Möller[1], Christian Griesinger[2], Teresa Carlomagno[3], Thomas E Exner[1,4]
[1]Department of Chemistry, University Konstanz, 78457 Konstanz, Germany; [2]Department of NMR-Based Structural Biology, MPIbpc, 37077 Göttingen, Germany; [3]Structural and Computational Biology Unit, EMBL, 69117 Heidelberg, Germany; [4]Zukunftskolleg, University Konstanz, 78457 Konstanz, Germany
E-mail: tim.tenbrink@uni-konstanz.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P39

Today's scoring functions are one of the main reasons that state-of-the-art protein-ligand dockings fail in about 20 % to 40 % of the targets due to the sometimes severe approximations they make. However these approximations are necessary for performance reasons. One possibility to overcome these problems is the inclusion of additional, preferably experimental information in the docking process. Especially ligand-based NMR experiments that are far less demanding than the solution of the whole complex structure are helpful.
Here we present the inclusion of three different types of NMR-data into the ChemPLP [1] scoring function of our docking tool PLANTS [2]. First, STD and intra-ligand trNOE spectra were used to obtain distant

constraints between ligand and protein atoms. This approach proved beneficial for the docking of larger peptide ligands i. e. the epitope of MUC-1 glycoprotein to the SM3 antibody [3].
In the second part the usefulness of INPHARMA data [4,5] is shown by combinig a score, evaluating the agreement between simulated and measured INPHARMA spectra, with the PLANTS ChemPLP scoring function. First results from rescoring after local optimization of the poses and full docking experiments are shown.

References

1. Korb O, Stützle T, Exner TE: *J Chem Inf Model* 2009, **49**:84-96.
2. Korb O, Stützle T, Exner TE: *Swarm Intell* 2007, **1**:115-134.
3. Korb O, Möller HM, Exner TE: *ChemMedChem* 2010, **5**:1001-1006.
4. Orts J, Tuma J, Reese M, Grimm SK, Monecke P, Bartoschek S, Schier A, Wendt KU, Griesinger C, Carlomagno T: *Angew Chem Int Ed* 2008, **47**:7736-7740.
5. Sanchez-Pedregal VM, Reese M, Meiler J, Blommers MJJ, Griesinger C, Carlomagno T: *Angew Chem Int Ed* 2005, **44**:4172-4175.

P40

Probabilistic classifier: generated using randomised sub-sampling of the feature space

Jonathan D Tyzack*, Hamse Y Mussa, Robert C Glen
Unilever Centre for Molecular Sciences Informatics, Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK
E-mail: jdt42@cam.ac.uk
Journal of Cheminformatics 2012, **4(Suppl 1):**P40

Nowadays supervised classification, based on the concept of pattern recognition, is an integral part of virtual screening. The central idea of supervised classification in chemoinformatics is to design a classifying algorithm that accurately assigns a new molecule to one of a set of predefined classes.
Naturally, probabilistic classifiers can be far more useful than hard point classifiers in making a decision on problems [1], such as virtual screening, where there is an associated risk in classifying an instance to one class or the other.
For their conceptual simplicity and computational efficiency probabilistic classification methods based on the Naive Bayes concept are widely employed in chemoinformatics. The simplicity of the Naive Bayes is due to the assumption that the descriptors representing the molecule one desires to classify are statistically independent. Unfortunately it is well documented that when the molecular descriptors are binary-valued - which is often the case in chemoinformatics - and thus take values of 0 or 1 the Naive Bayesian classifier can only act as a linear classifier in the descriptor space.
Techniques such as the Parzen-Window approach can address the above shortcomings but suffer from being computationally expensive as they require one to retain all the training dataset in core memory [2,3].
In an attempt to address the above mentioned drawbacks, a new probabilistic classifier is proposed which uses randomized sub-sampling of the descriptor space. The proposed algorithm generates better class membership predictions than its Naive Bayesian counterpart on classifying molecules that are non-linearly separable in descriptor space.
We present a realistic test of the new method by classifying large chemical datasets generated from the ChEMBL database [4].

References

1. Duda RO, Hart PE: **Pattern Classification and Scene Analysis.** John Wiley & Sons, Ltd : New York, NY 1973.
2. Parzen E: **The Annals of Mathematical Statistics.** 1962, **33**:1065-1076.
3. Harper G, Bradshaw J, Gittins JC, Green DVS, Leach AR: **.** *J Chem Inf Comput Sci* 2001, **41**:1295-1300.
4. **ChEMBL.** *J Comput-Aided Mol Des* 2009, **4**:195-198.

P41

Sodium dependent glucose transporter (SGLT) 1 / 2 - elucidating inhibitor SAR and selectivity using homology modelling and 3D QSAR studies

Susann Vorberg[1,2*], Ina Koch[2], Christian Buning[1]
[1]Sanofi-Aventis Deutschland GmbH, Industriepark Höchst, Building G838, 65926 Frankfurt a.M., Germany; [2]University Frankfurt, Institute for Computer Science, Department for Molecular Bioinformatics, 60325 Frankfurt a.M., Germany
E-mail: Susann.Vorberg@gmx.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P41

Inhibiting sodium-dependent glucose transporters (SGLTs) has been proposed as a new therapy for the treatment of diabetes [1]. SGLT2 as the most prominent member of this family is mainly expressed in the kidney and responsible for the reabsorption of the vast majority of the filtered glucose. This key role in the blood glucose homeostasis makes SGLT2 a promising target which has been clearly underlined by the results of preclinical and clinical studies. Therapeutic goals of SGLT2 inhibition are reduced plasma glucose levels and weight loss. In conjunction with the therapeutic benefits fewer side effects are expected than observed with other known diabetes drugs. Potential side effects in case of SGLT2 inhibition are expected to be mediated by a lack of selectivity towards SGLT1 that is mainly expressed in the intestine and responsible for glucose- and galactose absorption from food sources. Therefore, inhibition of SGLT1 leads to glucose- and galactose malabsorption, dehydration, and diarrhoea.
Consequently, any drug discovery project aiming for promising SGLT2 inhibitors has to take into account significant selectivity towards SGLT1. In order to describe the selectivity on a structural basis we extensively used molecular modelling techniques since no x-ray structural data is available for neither SGLT1 nor SGLT2. For both transporters homology models were generated using the published x-ray structure from vSGLT (PDB-code 3DH4) [2]. Structure based alignments of published SGLT2 inhibitors including selectivity data for SGLT1 were followed by 3D-QSAR studies utilizing CoMFA fields. Given a detailed analysis of the inhibitor SAR for both transporters the two structural models of SGLT1 and SGLT2 were compared leading to the identification of relevant differences and selectivity hot spots. In the future, the presented models could serve as a basis for the identification of new potent SGLT2 inhibitors that are selective towards SGLT1, too.

References

1. Chao E, Henry RR: *Nat Rev Drug Disc* 2010, **9**:551-559.
2. Faham S, *et al*: *Science* 2008, **321**:810-814.

P42

Postprocessing of molecular docking poses using binding free energy calculations

Kanin Wichapong*, Wolfgang Sippl
Department of Pharmaceutical Chemistry, Martin-Luther University Halle-Wittenberg, 06120, Halle(Saale), Germany
E-mail: kanin.wichapong@pharmazie.uni-halle.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P42

The main key to success in structure-based drug discovery is the accurate prediction of binding affinities of hit compounds. Molecular docking and scoring functions are often used for this purpose. However, it is often found that the top-ranked docking poses do not represent the right binding mode, and frequently there is no correlation between docking score and biological data. Therefore, "post-processing" of docking poses has recently got attraction. In previous work [1], we have successfully computed binding free energies (MM/PBSA) of 222 Wee1 kinase inhibitors and used the derived validated models for virtual screening. In the current work, we extended our studies to a data set of PIM1 kinase inhibitors. Cross-docking studies showed that the correct binding mode of the inhibitors can be determined after applying a post-processing procedure. The top-ranked docking poses gave wrong binding mode (high RMSD values ~4.0 Å), whereas the top-ranked poses selected after postprocessing yielded RMSD values around 0.5 Å. Subsequently, the docking poses giving the lowest binding free energy were selected and these values were used to establish a correlation with experimental data. A significant correlation between $\Delta G_{cal.}$ and ΔG_{exp} (r^2 = 0.58) was obtained. To summarize, the protocol described in the current work can be used for postprocessing of protein-ligand docking poses and for predicting biological activities of novel hits. Therefore, the protocol can be applied for structure-based optimization of hit molecules.

Reference

1. Wichapong K, Lawson M, Pianwanit S, Kokpol S, Sippl W: **Postprocessing of Protein-Ligand Docking Poses Using Linear Response MM-PB/SA: Application to Wee1 Kinase Inhibitors.** *J Chem Inf Model* 2010, **50**:1574-1588.

P43

MyChemise: a 2D drawing software that uses morphing for visualisation purposes

Jörg-Hubertus Wilhelm
Sophie-Scholl-Ring 26, 31275 Lehrte, Germany
E-mail: drjoerg-hubertus.wilhelm@arcor.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P43

2D drawing programs account for some of the first computer applications in chemistry and are widely distributed. Many publications, especially in recent time, show that this sector is developing continually [1-3]. MyChemise (My Chemical Structure Editor) is a new 2D structure editor [4]. It is designed as a Java applet that enables the direct creation of structures in the Internet using a web browser. MyChemise saves files in a digital format (.cse) and the import and export of .mol files using the appropriate connection tables is also possible.

MyChemise is available as a free online version in English and German from the author's homepage [5]. In addition to the known ways of drawing chemical structure formulas, there are also parts implemented in the program that allow the creation of different types of presentation. The morphing module uses this technology as a component for dynamic visualization. Four-point mapping (projective mapping) is mathematically solved using the unit squares method [6]. For example, it enables a clear and simple illustration of molecule vibrations and reaction sequences.

References

1. Ertl P: **Molecular structure input on the web.** *J Cheminf* 2010, **2**:1.
2. Dallakian P, Haider N: **FlaME: Flash Molecular Editor - a 2D structure input tool for the web.** *J Cheminf* 2011, **3**:6.
3. Kuhn S: **Open Source: Strukturen zeichnen.** *Nachr Chem* 2010, **58**:661-663.
4. Wilhelm JH: **MyChemise: A 2D drawing program that uses morphing for visualisation purposes.** *J Cheminf* 2011, **3**:53.
5. Authors homepage: http://www.knalltundstinkt.de.
6. Burger W, Burge MJ: **Digitale Bildverarbeitung.** Berlin: Springer-Verlag 2005, 377-385.

P44

INSARA: a new method for the analysis and visualization of Structure-Activity-Relationships

Sabrina Wollenhaupt*, Knut Baumann
Institut für Medizinische und Pharmazeutische Chemie, Technische Universität Braunschweig, Beethovenstr. 55, D-38106 Braunschweig, Germany
E-mail: s.wollenhaupt@tu-bs.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P44

Due to the rapid progress in combinatorial chemistry and (high-throughput-) screening, the organization and mining of the large amount of produced data becomes an increasingly important task in the modern drug discovery process. Herein, one particular challenge is the recognition of SAR-patterns e.g. for the selection of promising compounds for further analysis or lead optimization.

To support the medicinal chemist in doing this job a readily interpretable concept is required. Most published approaches addressing this problem (e.g. SARANEA [1]) use fingerprint similarity for the analysis of molecular relationships. Yet, a promising alternative and more intuitive way of comparing similarity is the maximum common substructure (MCS), the largest substructure in common between two molecules. Since computing the MCS is very demanding, it is usually not applicable to large data sets.

To circumvent this and other drawbacks (e.g. the exact match or incomplete ring problems) our own in-house strategy **INSARA** (**i**ntuitive **n**etworks for **S**tructure-**A**ctivity-**R**elationships **a**nalysis) employs reduced graphs (RG) instead of the original molecules in order to reduce the complexity of the problem to a manageable size.

The advantage of RG is that only pharmacophoric features and functional units represented by a few pseudoatoms have to be compared [2]. Iterative super- and substructure searches and MCS calculations subsequently lead in an unsupervised manner to SAR networks. When associated with bioactivity data the networks can be used for SAR analysis. While focussing on pharmacophoric properties a more general overview about the similarities within the active set is expected. For initial performance evaluation two targets with well-known SARs (ACE and COX 2) were chosen.

References

1. Lounkine E, Wawer M, Wassermann AM, Bajorath J: **SARANEA: A Freely Available Program to Mine Structure-Activity and Structure-Selectivity Relationship Information in Compound Data Sets.** *J Chem Inf Model* 2010, **50**:68-78.
2. Gardiner EJ, Gillet VJ, Willett P, Cosgrove DA: **Representing clusters using a maximum common edge substructure algorithm applied to reduced graphs and molecular graphs.** *J Chem Inf Model* 2007, **47**:354-366.

P45

In silico identification of novel PKC βII inhibitors: ligand and receptor based pharmacophore modeling, virtual screening, and molecular dynamics study

Baljinder K Grewal*, ME Sobhia
Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S.A.S Nagar, Punjab 160062, India
Journal of Cheminformatics 2012, **4(Suppl 1):**P45

The Protein Kinase C βII (PKCβII) belongs to conventional class of Protein kinase C enzyme and is preferentially activated during diabetic cardiomyopathy [1]. An effective inhibition of PKCβII is the potential option to directly treat the diabetic cardiomyopathy. Till date only one selective PKCβII inhibitor, ruboxistaurin reached phase III clinical trial for diabetic complications. Thus, there is an urgent need for exploring available chemical space for new PKCβII inhibitors. The sequential virtual screening workflow based on ligand and receptor based query was followed to identify novel PKCβII inhibitors. Three different strategies were followed for developing the ligand based model by HipHop module implemented in Catalyst, using: (I) three active and six moderately active compounds; (II) 17 active compounds; (III) docked poses of the compounds used in strategy (II). Receptor based query was developed based on the cocrystallised crystal structure of PKCβII with 2-methylbisindolylmaleimide (2mBIM) using the Unity module of Sybyl7.1. The best hypotheses from both methods consist of six features *viz.* one hydrogen bond donor (D), two hydrogen bond acceptor (A), two hydrophobic-aromatic (HYD) and one ring aromatic (R). Virtual screening scheme based on these 3D hypotheses identified a few molecules with higher docking score than the existing inhibitors. In addition, comparative molecular dynamics (MD) simulation studies of uncomplexed PKCβII and its complexes with 2mBIM, ruboxistaurin and newly identified compounds were performed to analyze the binding mode of the molecules. This study showed that complexed form of PKCβII was more stable than uncomplexed one during simulation period, and showed the stable H-bond formation with Glu421, Val423. This reveals the favorable interactions of identified compounds with PKCβII.

Reference

1. Murray NR, Baumgardner GP, Burns DJ, Fields AP: **Protein kinase C isotypes in human erythroleukemia (K562) cell proliferation and differentiation. Evidence that beta II protein kinase C is required for proliferation.** *J Biol Chem* 1993, **21**:15847-15853.

P46

Docking based 3D-QSAR studies applied at the BRAF inhibitors to understand the binding mechanism

Uzma Mahmood*, Zaheer ul-Haq
Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan
E-mail: mehmoodchemist@gmail.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P46

B-RaF is a member of the RAF family of serine/threonine kinases involved in the regulation of cell growth, differentiation, and proliferation. It forms part of a conserved apoptosis signals through the RAS-RAF-MAPK pathway. As an important target of the cancer treatment B-Raf has more potential for researcher. The discovery of the ^{V600E}B-RAF mutation [1] has elevated expectation for targeted therapy against human melanoma. In the current work the molecular modeling study were carried out very first time with 3D-QSAR studies [2] by correlating the docking protocol for three different datasets of B-Raf inhibitors. Based on the co-crystallized compound (PDB

ID: 1UWJ), several alignment methods were utilized to derive reliable CoMFA and CoMSIA models. The best CoMFA model (q^2 = 0.753, r^2 = 0.807). With the same alignment protocol, a statistically reliable CoMSIA model # 14 was also derived (q^2 = 0.962, r^2 = 0.961). The actual predictive powers of both models were thoroughly validated with an external test set, which gave satisfactory predictive r^2 values of 0.89 and 0.90, respectively. Contour maps from CoMFA and CoMSIA models, supporting the statistical results and revealed important structural features responsible for increasing biological activity within the active site and explained the correlation between biological activity and receptor-ligand interactions. These results can offer useful information for the design of new B-Raf inhibitor.

References

1. Davies H, Bignell GR, Bottomley W: **Mutations of the BRAF Gene in Human Cancer.** *Nature* **417**:949-954.
2. Ul-Haq Z, Mahmood U, Jehangir B: **Ligand-based 3D-QSAR Studies of Physostigmine Analogues as Acetylcholinesterase Inhibitors.** *Chem Biol Drug Des* 2009, **74**:571-581.

P47

Preliminary characterization of N-trimethylchitosan as a nanocarrier for malarie vaccine

Petra O Nnamani[1,2*], Ngozi J Nwodo[1], Scoles Giacinto[2]
[1]Drug Discovery and Drug Delivery Research Unit, Faculty of Pharmaceutical Sciences, University of Nigeria, Nsukka 410001, Enugu State, Nigeria; [2]Nanotechnology Unit, ICS-UNIDO International Center for High Technology and New Materials, AREA Science Park, Padriciano 99, Trieste, Italy
Journal of Cheminformatics 2012, **4(Suppl 1):**P47

Vaccination is considered to be most effective way of fighting infectious diseases like malaria etc [1]. N, N, N-trimethylchitosan (TMC) was synthesized from chitosan. Nanoparticles of the TMC were prepared in various media (milliQ water, Na_2CO_3 (pH 10.92), Na_2HPO_4 (pH 9.01) and alhydrogel® beads which were characterized as adjuvant for possible vaccine delivery. The nanoparticles were analyzed using microscopy (Phase contrast microscope and Confocal laser scanning microscope), and Malvern zetasizer Nano- ZS. Time-resolved particle size analysis was performed after one month storage of the TMC nanoparticles at 4 °C. The result of the study showed that PBS was the best medium that produced cationic, monodispersed and stable TMC nanoparticles of less than 65 nm forming a compatibly homogeneous system even upon storage. Microscopy of the polyelectrolyte doped nanoparticles showed a clear coating due to PSS at the periphery of the particles and a fluorescent core with some tiny central hollow cavities Confocal microscopy of the alhydrogel beads showed particle size of 1.6 μm. The fluorescent dye (PSSRhodamine) coated the entire particle surface suggesting a more or less adsorption process for the antigen delivery [2]. Hence, the hope of nanocarrier for malaria vaccine.

References

1. Kammona O, *et al*: **Nanocarrier Aided Nasal Vaccination: An Experimental and Computational Approach.** *Ind Eng Chem Res* 2011, **50**:590-601.
2. Ajay P, *et al*: **A review on novel lipid based nanocarriers.** *Int J Pharm and Pharm Sci* 2010, **2**:30-35.

P48

Development of target focused library against drug target of *P. falciparum* using SVM and Molecular docking

Sangeetha Subramaniam[1*], Monica Mehrotra[2], Dinesh Gupta[1]
[1]Structural and Computational Biology, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi, 110070, India; [2]Department of Computer Science, Jamia Millia Islamia, New Delhi, 110025, India
E-mail: sangi@icgeb.res.in
Journal of Cheminformatics 2012, **4(Suppl 1):**P48

PfHslV, a homolog of β subunit of 20S proteasome forms the proteolytic core of the *PfHslUV* machinery in *P. falciparum* [1,2]. *PfHslV* has no homolog in the human host and it is a promising drug target essential to the plasmodial metabolism. The use of single proteasome inhibitor targeting these threonine proteases has a potential to be antimalarial drug candidate. One of our recent studies identified several promising inhibitors against 20S β5 subunit of *P. falciparum* [3]. The present study adopts a similar knowledge based virtual screening strategy using Support Vector Machines (SVM) and molecular docking to build a focused library of potential *PfHslV* inhibitors. SVM model has been trained using 170 molecular descriptors of 64 inhibitors and 208 putative non-inhibitors. The non-linear classifier based on Radial Basis Function (RBF) kernel yielded classification accuracy of 97%. The SVM model rapidly predicted inhibitors from NCI library and were subsequently docked in to the active site of an optimised three-dimensional model of *PfHslV*. The novel drug-like *PfHslV* inhibitors with very good binding affinity and novel scaffold can be a good starting point to develop new antimalarial drugs.

References

1. Ramasamy G, Gupta D, Mohmmed A, Chauhan VS: **Characterization and localization of Plasmodium falciparum homolog of prokaryotic ClpQ/HslV protease.** *Mol Biochem Parasitol* 2007, **152**:139-148.
2. Subramaniam S, Mohmmed A, Gupta D: **Molecular modeling studies of the interaction between Plasmodium falciparum HslU and HslV Subunits.** *J Biomol Struct Dyn* 2009, **26**:403-524.
3. Subramaniam S, Mehrotra M, Gupta D: **Support Vector Machine based prediction of P. falciparum proteasome inhibitors and development of focused library by molecular docking.** *Comb Chem High Throughput Screen* 2011, **14(10)**:898-907.

P49

The GMX-Plugin for the CELLmicrocosmos MembraneEditor

Sebastian Rubert[1], Christian Gamroth[1], André J Heissmann[1], Gunther Lukat[1], Ralf Rotzoll[1], Alexander Schäfer[1], Jens Krüger[2], Björn Sommer[1*]
[1]Bio-/Medical Informatics Department, Bielefeld University, Bielefeld, NRW, 33615, Germany; [2]Department of Chemistry, University of Paderborn, Paderborn, NRW, 33098, Germany
E-mail: bjoern@CELLmicrocosmos.org
Journal of Cheminformatics 2012, **4(Suppl 1):**P49

Membrane research *in silico* can roughly be subdivided into three parts: the modeling, simulation and analysis process. The GMX-plugin tries to bridge the gap between these three parts represented by the tools CELLmicrocosmos MembraneEditor (CmME) and Gromacs (GMX).
CmME was developed to enable students and researchers a generation of PDB-based membranes in a fast and intuitive way without high computational requirements. From the beginning it was developed as an independent Javabased web start tool with a user-interface providing direct access to all functions implemented. The high performance of most Membrane Packing Algorithms is achieved especially by the handling of molecules as inflexible structures. The generated membranes can be exported to a PDB-file to be used with external applications [1].
One of these programs is GMX (here version 4.5.X), the well-known Molecular Dynamics package, supported and used over one decade by a very large community. It is applicable to the simulation of peptides, proteins, lipids as well as complete membranes [2].
The GMX-plugin version 1.1 is intended as an interface between CmME and GMX. It is combined with CmME on the local system and is able to access GMX on a local machine or on an external high-performance system via ssh or Unicore [3]. It is packaged with a set of lipids compatible to the Gromos 45a3 forcefield. In addition, predefined protocols exist for immediately starting a simulation of CmME-generated membranes. Custom protocols may be created, saved and reloaded by the user.
The beta version of the plugin can be downloaded at: http://Cm2.CELLmicrocosmos.org.

References

1. Sommer B, Dingersen T, Gamroth C, Schneider S, Rubert S, Krüger J, Dietz K-J: **CELLmicrocosmos 2.2 MembraneEditor: A Modular Interactive Shape-Based Software Approach To Solve Heterogeneous Membrane Packing Problems.** *J Chem Inf Model* 2011, **51**:1165-1182.
2. **Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.** *J Chem Theory Comput* 2008, **4**:435-447.
3. Schuller B, Schumacher M: **Space-Based Approach to High-Throughput Computations in UNICORE 6 Grids.** *Lect Notes Comput Sc* 2009, **5415**:75-83.

P50

Open access: changing the way chemistry is published

Jan Kuras*, Bailey Fallon
Chemistry Central, London WC1X 8HL, UK
E-mail: jan.kuras@chemistrycentral.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P50

Publishing research in open access (OA) journals ensures free and permanent unrestricted online access to peer-reviewed articles [1], and increased visibility for articles and increased potential citations. In addition, authors retain copyright to their work, and data can be redistributed, reused and translated freely [2]. Removing barriers to accessing and sharing data is beneficial for researchers and their institutions, as well as funders, educators and the public.

Having had success in biomedical and physics publishing, support continues to grow for open access in chemistry, in particular amongst the cheminformatics community. The range of benefits for authors and institutions makes open access a viable and effective publishing model for chemistry research.

This poster provides an overview of the benefits of open access, and compares the OA model to traditional subscription journal publishing.

References

1. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 2000 [http://oa.mpg.de/files/2010/04/berlin_declaration.pdf], T. Others & V Clever.
2. Creative Commons copyright license. [http://creativecommons.org/licenses/].

P51

How often do follow-on activities occur - trends seen in a patent database for GPCRs

Christian Tyrchan[1*], Sorel Muresan[2]
[1]AstraZeneca R&D, CVGI iMed, Pepparedsleden 1, S-43183 Mölndal, Sweden; [2]AstraZeneca R&D, DECS, Pepparedsleden 1, S-43183 Mölndal, Sweden
E-mail: christian.tyrchan@astrazeneca.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P51

The first compound reaching the market following the approval of the official offices is referred to as first-in-class or breakthrough drug. These compounds constitute a new class of drugs meaning that they introduce a novel mode of action (MoA) or provide a significant improvement over the standard therapy, if not enables a therapy, in terms of efficacy and safety. A follow-on (or me-too) drug is, in the common understanding, a chemical entity which has structural similarity or has the same pharmacological MoA as the first-in-class drug. The approach of follow-on drugs is controversially discussed in the literature [1-3].

Based on the GVKBIO Medicinal Chemistry and Target Class databases (which capture explicit relationships between published documents, compounds, assay results and targets), we investigated the occurrence of drug discovery follow-on activities as captured by the pharmaceutical patent space [4]. To do so, we abstracted from the GVKBIO databases 11,827 patents which are linked to a GPCR target with a defined Entrez Gene ID (as November 2010). In a next step, we removed all peptide structures because of the size and the similarity of their backbone. Subsequently, we kept all 10,253 patents belonging to the TOP 100 companies in terms of number of patents published. The set was consolidated by merging patents for known mergers and acquisitions until 2008. All possible patent combinations were created if they share an official gene name and are published by different companies within a 6-year interval. Similarity descriptors and scores could be generated for 1,570,381 out of a total of 1,608,368 pairs (97.6%) [5].

As expected, the pharmaceutical research is highly competitive and dynamic. This is supported by the fact that most of the follow-on patents are published within the first two years with a high number of patent pairs published in the same year. Around 4800 (47%) patents included in our analysis are linked to follow-on activities with only small differences between the companies.

References

1. Cohen J, Cabanilla L, Sosnov J: Role of follow-on drugs and indications on the WHO Essential Drug List. *J Clin Pharm Ther* 2006, **6**:585-592.
2. Dimasi JA, Faden LB: Competitiveness in follow-on drug R&D: a race or imitation? *Nat Rev Drug Discov* 2011, **1**:23-27.
3. Hollis A: Comment on "The economics of follow-on drug research and development: trends in entry rates and the timing of development". *Pharmacoeconomics* 2005, **12**:1187-92.
4. [http://www.gvkbio.com/informatics.html].
5. Godden JW, Bajorath J: Differential Shannon Entropy as a sensitive measure of differences in database variability of molecular descriptors. *J Chem Inf Comput Sci* 2001, **4**:1060-1066.

P52

From chemical shift data through prediction to assignment and NMR LIMS - multiple functionalities of nmrshiftdb2

Stefan Kuhn[1*], Nils E Schlörer[2], Heinz Kolshorn[3], Raphael Stoll[4]
[1]Aureliusstr. 45, 52067 Aachen, Germany; [2]Department of Chemistry, University of Cologne, 50939 Köln, Germany; [3]Institut für Organische Chemie, Johannes-Gutenberg-Universität, 55128 Mainz, Germany; [4]Biomolecular NMR, Ruhr-Universität Bochum, Bochum, Germany
E-mail: stefhk3@web.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P52

nmrshiftdb2 and its predecessor NMRShiftDB [1,2] have been available as a community-based NMR database since 2002. During that time a continuously growing set of currently more than 40000 structures with 48600 spectra could be established. These data are freely available (http://nmrshiftdb.org) and cannot only be searched but can also be downloaded and used for scientific investigations. Supplementary to the database, supplementary software was developed including an NMR lab administration system.

Recently, there were some changes in the team and we had a rebranding to nmrshiftdb2. Now, we would be very interested to enter discussion about the project and its future perspectives by receiving feedback from former, current and potential users of different areas. We will discuss the state of this project, e.g. software functions, data collection, and use for research.

An important new feature is the laboratory information management system (LIMS) which has been developed with the intention to better integrate the database functionality into an academic NMR laboratory environment. Nmrshiftdb2 now allows NMR laboratories at the same time to administer and account for their users, instruments and measurements and use the known database functions from the same surface.

References

1. Steinbeck C, Kuhn S: NMRShiftDBConstructing a Free Chemical Information System with Open-Source Components. *Phytochemistry* 2004, **65**:2711-2717.
2. Steinbeck C, Krause S, Kuhn S: NMRShiftDB – compound identification and structure elucidation support through a free community-built web database. *J Chem Inf Comput Sci* 2003, **43**:1733-1739.

P53

A computational model for predicting the transport of compounds by ABCC2

Marta Pinto*, Gerhard F Ecker
Department of Medicinal Chemistry, University of Vienna, Vienna, 1090, Austria
E-mail: marta.pinto@univie.ac.at
Journal of Cheminformatics 2012, **4(Suppl 1):**P53

The ATP-binding cassette (ABC) proteins represent a large family of transmembrane proteins that use the energy of ATP hydrolysis to transport a wide variety of physiological substrates across biological membranes [1]. Of them, particular attention has been focused in the last years on the role of the ABCC2 transporter in drug clearance and disposition.

The ABCC2 transporter is a transmembrane protein expressed in the apical cell membrane of hepatocytes and epithelial cells of small intestine and kidney, where it is involved in the elimination of many endogenous and exogenous substrates from the cell, including compounds clinically relevant [2]. Alteration in the disposition and elimination of these compounds can modify their pharmacokinetic and pharmacological profiles, leading to reduced efficacy or increased toxicity.

In this scenario, the aim of the present work was the development of an *in silico* model based on the Gottesman database [3] able to predict if certain compounds of interest are ABCC2 substrates or not. To this end, several machine learning methods were explored using the WEKA data mining software [4]. Molecules were represented by 2D and 3D descriptors calculated with the MOE software [5]. Feature selection was used to improve the efficiency of the data mining algorithms and identify the contribution of different features. Misclassification cost was used in order to deal with data set imbalance. According to our results, naive Bayesian updatable (NBU) had the highest performance, with an overall prediction accuracy of 72.1% and a Matthew's correlation coefficient of 0.44. Furthermore, sensitivity and specificity values were significantly improved with values of 72.7% and 71.6%, respectively.

References

1. Higgins CF: **ABC transporters: from microorganisms to man.** *Annu Rev Cell Biol* 1992, **18**:67-113.
2. Marquez B, Van Bambeke F: **ABC Multidrug Transporters: Target for Modulation of Drug Pharmacokinetics and Drug-Drug Interactions.** *Curr Drug Targets* 2011, **12**:600-620.
3. Szakács G, Annereau JP, Lababidi S, Shankavaram U, Arciello A, Bussey KJ, Reinhold W, Guo Y, Kruh GD, Reimers M, Weinstein JN, Gottesman MM: **Predicting drug sensitivity and resistance: Profiling ABC transporter genes in cancer cells.** *Cancer Cell* 2004, **6**:129-137.
4. [http://www.cs.waikato.ac.nz/ml/].
5. [http://www.chemcomp.com/].

P54

Insights into binding events of GABA- and Tiagabine- analogues in the Gamma-Aminobutyric Acid Transporter 1 by means of Molecular Modelling

Barbara Zdrazil[1*], Andreas Jurik[1], Regina Reicherstorfer[1], Thomas Stockner[2], Harald H Sitte[2], Gerhard F Ecker[1]
[1]University of Vienna, Department of Medicinal Chemistry, Pharmacoinformatics Research Group, Althanstrasse 14, A-1090 Vienna, Austria; [2]Medical University Vienna, Institute of Pharmacology, Währinger Straße 13A, A- 1090 Vienna, Austria
E-mail: barbara.zdrazil@univie.ac.at
Journal of Cheminformatics 2012, **4(Suppl 1):**P54

The human transporters for the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) hGAT-1, 2, and 3, and hBGT-1 belong to the neurotransmitter-sodium symporter (NSS) family of membrane transport proteins. hGAT-1 has been a target for the design of antiepileptic therapeutics [1], with tiagabine (Gabitril®) being the only GAT inhibitor on the market. The lack of specific inhibitors for the other hGAT subtypes, results from a still missing detailed understanding of the molecular basis of drug-transporter interactions of the respective subtypes.

We aim at elucidating plausible binding modes for ligands of each subtype, respectively. In our first studies, we built up homology models for hGAT-1 in the occluded and outward-facing conformation based on the respective high resolution structures of the leucine transporter of Aquifex aeolicus (LeuT) (pdb-codes: 2A65 and 3F3A). Afterwards, the natural substrate GABA was docked into the occluded state model and tiagabine into the outward-facing model by making use of the Induced Fit Docking module of Schrödinger, LLC. Both models were further subject to extensive Molecular Dynamics (MD) simulation studies (GROMACS 4.5.3 was used).

By the aid of MD simulations we could detect the existence of conserved water molecules into the GABA (occluded) and tiagabine (open-to-out) binding sites, respectively. Average structures from the equilibrated trajectories were extracted and subsequently served for further docking experiments.

Docking of small ligands (GABA, Guvacine and R-/S Nipecotic Acid) into the occluded state model nicely indicated their preference to bind in an extended conformation (also demonstrated by long-term MD with GABA). MD and Docking of tiagabine and analogues into the open-to-out state model revealed a common binding mode.

Additionally, our studies will be extended to the other hGAT subtypes so as to elucidate the secret of subtype selectivity of GABA Transporters.

Acknowledgements: We acknowledge financial support provided by the Austrian Science Fund, grant F3502 and grant F3506.

Reference

1. Palló A, Simon A, Bencsura A, Héja L, Kardos J: **Substrate-Na+ complex formation: coupling mechanism for gamma-aminobutyrate symporters.** *Biochem Biophy Res Commun* 2009, **385**:210-214.

P55

In silico pK_a prediction

Robert Körner[1,2*], Iurii Sushko[1], Sergii Novotarskyi[1,2], Igor V Tetko[1,2]
[1]eADMET GmbH, Munich, 85764, Germany; [2]Helmholtz Zentrum München, Munich, 85764, Germany
E-mail: r.koerner@eadmet.com
Journal of Cheminformatics 2012, **4(Suppl 1):**P55

The biopharmaceutical profile of a compound depends directly on the dissociation constants of its acidic and basic groups, commonly expressed as the negative decadic logarithm pK_a of the acid dissociation constant (K_a). The acid dissociation constant (also protonation or ionization constant) K_a is an equilibrium constant defined as the ratio of the protonated and the deprotonated form of a compound. The pK_a value of a compound strongly influences its pharmacokinetic and biochemical properties. Its accurate estimation is therefore of great interest in areas such as biochemistry, medicinal chemistry, pharmaceutical chemistry, and drug development. Aside from the pharmaceutical industry, it also has relevance in environmental ecotoxicology, as well as the agrochemicals and specialty chemicals industries.

In literature, a vast number of different approaches for pK_a prediction can be found [1]. These approaches can be divided into two different classes. On the one hand there are direct calculations, so called ab initio methods, trying to determine the pK_a value by quantum chemical or mechanical computation. On the other hand, statistical models, trained on chemical or structural descriptors. These descriptors can be, for example, of quantum chemical, semi empirical, graph topological or simple statistical nature. This type of modeling is called QSPR (Quantitative Structure Property Relationship).

In our recent work, we develop such a QSPR model using localized molecular descriptors to train multiple linear regression and artificial neural networks to estimate dissociation constants (pK_a). The performance of our approach is similar to that of a semi-empirical model based on frontier electron theory [2] as well as a prediction model based on Graph Kernels [3].

How such a prediction model can be built, is shown by an example performed with OCHEM, an *online chemical database with an environment for modeling* (http://ochem.eu/). It is a publicly accessible database for chemical compound data and predictive models. Further, users get the facility to develop, apply, and distribute predictive models, so it is unique in its combination of compound data and predictive models.

References

1. Rupp M, Körner R, Tetko IV: **Predicting the pKa of small molecule. Combinatorial chemistry & high throughput screening.** 2011, **14(5)**:307-327.
2. Tehan BG, *et al*: **Estimation of pKa Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids.** *Quant Struct-Act Relat* 2002, **21**:457-472.
3. Rupp M, Körner R, Tetko IV: **Estimation of Acid Dissociation Constants Using Graph Kernels.** *Molecular Informatics* 2010, **29**:731-740.

P56

Performance of dispersion-corrected density functional theory for thermochemistry and non-covalent interactions

Waldemar Hujo*, Stefan Grimme
Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, D-53115 Bonn, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P56

The accuracy of non-local van der Waals density functional [1] is tested for the thermochemical properties of 1200+ atoms and molecules in the GMTKN30 database. Five (hybrid)GGA functionals are augmented by the non-local (NL) part of the VV10 functional. The widely used atom-pair wise dispersion correction DFT-D3 [2] is considered for comparison. The

addition of the NL dispersion energy definitely improves the results of all tested short-range functionals. Based on little empiricism and basic physical insight, DFT-NL can be recommended as robust electronic structure method.

For more detailed insight into non-covalent bonding, potential energy curves [3] for five complexes with weak to medium strong hydrogen bonds have been computed with dispersion corrected DFT methods VV10, DFT-D3 and vdW-DF2 [4]. All dispersion corrected methods perform reasonably well for these hydrogen bonds. For the fluorinated complexes, the VV10 method gives remarkably good results. The vdW-DF2 method yields good interaction energies similar to the other methods, but fails to provide accurate equilibrium separations. For large-scale applications we can recommend DFT-D3 based structure optimizations with subsequent checking of interaction energies by single-point VV10 computations.

References

1. Vydrov OA, van Voorhis T: . *J Chem Phys* 2010, **133**:244103.
2. Grimme S, Antony J, Ehrlich S, Krieg H: . *J Chem Phys* 2010, **132**:154104.
3. Hujo W, Grimme S: . *Phys Chem Chem Phys* 2011, **13**:13942.
4. Lee K, Murray ED, Kong L, Lundqvist Bl, Langreth DC: . *Phys Rev B* 2010, **82**:081101.

P57

Fragment-based identification of multi-target ligands by self-organizing map alignment

Janosch Achenbach*, Franca-Maria Klingler, Steffen Hahn, Svenja Steinbrink, Mirjam Schroeder, Frank Loehr, Volker Doetsch, Dieter Steinhilber, Ewgenij Proschak
Institute of Pharmaceutical Chemistry, LiFF/OSF/ZAFES, Goethe University, Frankfurt am Main, Germany, Max-von-Laue Str.9, D-60438, Frankfurt/M, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P57

In the recent years the prevalent paradigm in drug discovery of „one drug – one target – one disease", following the assumption that highly selective ligands would avoid unwandted side effects caused by binding to seconday non-theratpeutic targets, got reconsidered. The results of post-genomic and network biology showed that proteins rarely act in isolated systems but rather as a part of a highly connected network [1]. It was further shown that the efficacy of several approved drugs is traced back to the fact that they act on multiple targets [2]. Therefore inhibiting a single target of such a network might not lead to the desired therapeutic effect. These findings lead to a shift towards polypharmacology [3] and the rational design of selective multi-target drugs, which have often improved efficacy [4]. But the design of multi target drugs is still a great challenge in regard of a sufficient activity on each target as well as an adequate pharmacokinetic profile [5]. Early design strategies tried to link the pharmacophors of known inhibitors, however these methods often lead to high molecular weight and low ligand efficiency.

We present a new approach based on self-organizing maps [3,6] (SOM) for the identification of multi-target fragments. We describe a workflow that initially identifies multi-target relevant substructures with a combination of maximum common substructure search and the alignment of multiple SOMs. Furthermore, these substructures are trained together with a fragment library on additional SOMs to find new multi-target fragments, validated by saturation transfer difference (STD)-NMR and biochemical assay systems. We used our approach for the identification of new dual-acting inhibitors of 5-Lipoxygenase (5-LO) and soluble Epoxide Hydrolase (sEH), both enzymes located in the arachidonic acid cascade and involved in inflammatory processes, pain and cadiovascular diseases.

References

1. Jeong HM, Barabási A-L, Oltvai ZN: *Nature* 2001, **411**:41-42.
2. Yildirim Ma, Goh K-I, Cusick ME, Barabási A-L, Vidal M: *Nature Biotechnology* 2007, **25**:1119-1126.
3. Achenbach J, Tiikkainen P, Franke L, Proschak E: **Future medicinal chemistry.** 2011, **3**:961-968.
4. Morphy R, Rankovic Z: *J Med Chem* 2005, **48**:6523-6543.
5. Morphy R, Kay C, Rankovic Z: *Drug discovery today* 2004, **9**:641-651.
6. Kohonen T: **Biological Cybernetics.** 1982, **43**:59-69.

P58

Discovery of novel TLR modulators by Molecular Modeling and Virtual Screening

Manuela S Murgueitio[1*], Sandra Santos-Sierra[2], Gerhard Wolber[1]
[1]Pharmaceutical Chemistry, Institute of Pharmacy, Freie Universität Berlin, Berlin, Berlin,14195, Germany; [2]Institute of Biochemical Pharmacology, Medizinische Universität Innsbruck, A-6020, Austria
E-mail: m.murgueitio@fu-berlin.de
Journal of Cheminformatics 2012, **4(Suppl 1):**P58

Toll-like receptors (TLRs) play a crucial role in the onset of innate immunity by distinguishing between endogenous and pathogen-associated molecular patterns. TLR2, in cooperation with TLR1 and TLR6, recognizes several microbial components such as lipoteichoic acids and lipoproteins [1]. Toll-like receptors have been broadly reported to contribute to several inflammatory chronic diseases and autoimmune diseases [2]. In this study we aim to discover new TLR2 modulating agents through computer-aided drug design. Based on recently identified synthetic TLR2 agonists [3] and antagonists [4], a shape and chemical-feature based similarity search was performed against a library of 260.071 compounds provided by the National Cancer Institute (NCI) [5]. This led to several virtual hits, which were tested *in vitro* in a cell-based assay. Several compounds with biological activity on TLR2 signaling in general and TLR1 signaling specifically were identified.

To further optimize these biologically validated virtual hits, molecular interaction fields (MIFs) for the dimerization of TLR2 and TLR1 were developed. Feature-based MIFs allowed for the manual creation of virtual compounds that fulfill an optimized interaction pattern, which led to a 3D pharmacophore that was used for a second virtual screening to select compounds for biological testing.

References

1. Akira S, Takeda K: **Toll-like Receptor Signalling.** *Nat Rev Immunol* 2004, **4(7)**:499-511.
2. Kawai T, Akira S: **The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors.** *Nat Rev Immunol* 2010, **11(5)**:373-384.
3. Guan Y, *et al*: **Identification of Novel Synthetic Toll-like Receptor 2 Agonists by High Throughput Screening.** *J Biol Chem* 2010, **258(31)**:23755-23762.
4. Zhou S, *et al*: **Discovery of a novel TLR2 signaling inhibitor with anti-viral activity.** *Antiviral Res* 2010, **87(3)**:295-306.
5. [http://cactus.nci.nih.gov/download/nci/].

P59

SiteBinder – an improved approach for comparing multiple protein structural motifs. Case studies on biologically important motifs

David Sehnal[1*], Radka Svobodová Vařeková[1], Heinrich J Huber[2], Stanislav Geidl[1], Crina-Maria Ionescu[1], Michaela Wimmerová[1], Jaroslav Koča[1]
[1]National Centre for Biomolecular Research and CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic, 625 00, CZ; [2]Systems Biology Group, Royal College of Surgeons in Ireland, Dublin, Ireland, D2, IE
E-mail: david.sehnal@mail.muni.cz
Journal of Cheminformatics 2012, **4(Suppl 1):**P59

Novel high-throughput experimental techniques produce a large amount of data on the 3D structure of proteins and their structural motifs. These motifs can be used as patterns in drug discovery [1], can help to understand the relationship between a protein's structure and its function [2] and to classify proteins [3]. In order to extract as much information as possible from this data, new techniques and tools are necessary, and among them fast approaches to perform the multiple superimposition of large sets of protein structural motifs. We report here on the development of such a tool.

We have implemented our newly developed multiple superimposition methodology in the web application SiteBinder, which is able to process hundreds of protein structural motifs in a very short time and provides an intuitive and user-friendly interface. We also demonstrate the applicability of SiteBinder using three case studies, focused on biologically important protein motifs. In the first case study, we compared the structures of 67 PA-IIL sugar binding sites containing 9 different sugars and we found that the sugar binding sites of PA-IIL and its mutants have a conserved structure despite their binding different sugars. The second case study focused on

more than 300 zinc finger central motifs and revealed that the molecular structure in the vicinity of the Zn atom in Cys2His2 zinc fingers is highly conserved. In the last case study, we superimposed 12 BH3 domains from pro-apoptotic proteins, and found that there is a structural basis for the functional segregation of these proteins into activators and enablers.

References

1. Baran I, Svobodová Vařeková R, Parthasarathi L, Suchomel S, Casey F, Shields DC: **Identification of potential small molecule peptidomimetics similar to motifs in proteins.** *J Chem Inf Model* 2007, **47**:464-474.
2. Watson JD, Laskowski RA, Thornton JM: **Predicting protein function from sequence and structural data.** *Curr Opin Struct Biol* 2005, **15**:275-284.
3. Chang YS, Gelfand TI, Kister AE, Gelfand IM: **New classication of supersecondary structures of sandwich-like proteins uncovers strict patterns of strand assemblage.** *Proteins* 2007, **68**:915-921.

P60

Searching for tunnels of proteins – comparison of approaches and available software tools

Deepti Jaiswal*, Radka Svobodová Vařeková, Crina-Maria Ionescu, David Sehnal, Jaroslav Koča
National Centre for Biomolecular Research and CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic, 625 00, CZ
Journal of Cheminformatics 2012, **4(Suppl 1):**P60

Tunnels are access paths connecting the interior of molecular systems with the surrounding environment. The presence of tunnels in proteins influences their reactivity, as they determine the nature and intensity of the interaction that these proteins can take part in. A few examples of systems whose function relies on tunnels include transmembrane proteins involved in small molecule transport and signal transduction, peptide exit channels through which ribosomes release newly synthesized proteins during transcription Knowledge of the location and characteristics of protein tunnels can find immediate applications in rational drug design, protein engineering, enzymology etc.

Identification and characterization of tunnels has been the focus of several studies, and various algorithms and software tools have been developed for these purposes [1-4]. These methodologies use special mathematical algorithms to represent and scan the surface of the protein in search for tunnels and the amino acid residues involved.

In the presented study we perform a benchmarking study of the most known approaches and software tools for finding tunnels in proteins (Mole, MolAxis, Hollow, etc.). We focused on proteins from the cytochrome P450 family, which are very important from the biological point of view. We provide a critical discussion of the strong and weak points of the analyzed approaches and software tools.

References

1. Petřek M, Kosinová P, Koča J, Otyepka M: **MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels.** *Structure* 2007, **15**:1357-1363.
2. Ho BK, Gruswitz F: **HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures.** *BMC Struct Biol* 2008, **8**:49.
3. Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov R: **MolAxis: efficient and accurate identification of channels in macromolecules.** *Proteins* 2008, **73**:72-86.
4. Petřek M, Otyepka M, Banáš P, Kosinová P, Koča J, Damborský J: **CAVER: a new tool to explore routes from protein clefts, pockets and cavities.** *BMC Bioinformatics* 2006, **7**:316.

P61

How the methodology of 3D structure preparation influences the quality of QSPR models?

Stanislav Geidl*, Roman Beránek, Radka Svobodová Vařeková, Tomáš Bouchal, Miroslav Brumovský, Michal Kudera, Ondřej Skřehota, Jaroslav Koča
National Centre for Biomolecular Research and CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic, 625 00, CZ
Journal of Cheminformatics 2012, **4(Suppl 1):**P61

QSPR modelling is a very useful and popular methodology for estimating the physical and chemical properties of molecules. The inputs for QSPR models are 3D structures of molecules. Currently, the 3D structures for millions of molecules are publicly available. A large number of these 3D structures were generated by software tools for the conversion of 2D structures into 3D. Moreover, the generated structures can be geometrically optimized by different approaches, such as molecular mechanics, quantum mechanics, etc. The question arises as to how the methodology of 3D structure preparation influences the quality of QSPR models that use these structures. Is there some software tool for 3D structure construction more suitable for QSPR modelling purposes than others? Conversely, which software tools are inappropriate? How strong is the influence of the geometry optimization procedure? We focused on these questions in the present study.

In our work, we analyzed the influence of 3D structure preparation methodology on the quality of QSPR models for the prediction of the acid dissociation constant (pK_a) from atomic charges [1]. We employed three different software tools for 3D structure generation (Corina [2], Balloon [3], etc.), together with two approaches for geometry optimization. This way, we prepared nine sets of 3D structures, and used them to develop QSPR models based on several different charge calculation schemes. Afterwards, we compared the accuracy of these QSPR models and discussed the influence of the methodology for 3D structure preparation.

References

1. Svobodová Vařeková R, Geidl S, Ionescu CM, Skřehota O, Kudera M, Sehnal D, Bouchal T, Abagyan R, Huber JH, Koča J: **Predicting pK_a values of substituted phenols from atomic charges.** *J Chem Inf Model* 2011, **51**:1795-1806.
2. Sadowski J, Gasteiger J: **From atoms and bonds to three-dimensional atomic coordinates: Automatic model builders.** *Chem Rev* 1993, **93**:2567-2581.
3. Vainio MJ, Johnson MS: **Generating conformer ensembles using a multiobjective genetic algorithm.** *J Chem Inf Model* 2007, **47**:2462-2474.

P62

QSAR modeling for In vitro assays: linking ToxCast™ database to the integrated modeling framework "OCHEM"

Ahmed Abdelaziz, Iurii Sushko, Wolfram Teetz, Robert Körner, Sergii Novotarskyi, Igor V Tetko*
Chemoinformatics group, Institute of Bioinformatics & Systems Biology, 85764 Neuherberg, Germany
Journal of Cheminformatics 2012, **4(Suppl 1):**P62

ToxCast™ project, phases I and II, is testing a combined total of 960 unique chemicals with more than 650 high-throughput assays. The aim of this database is to use advanced science tools to help understand how human body processes are impacted by exposures to chemicals and helps determine which exposures are most likely to lead to adverse health effects. To better serve this goal and to allow In silico analysis of In vitro assays, we linked the database with an integrated QSAR modeling framework.

The Online Chemical Modeling Environment is a web-based platform that aims to automate and simplify the typical steps required for QSAR modeling. The platform consists of two major subsystems: the database of experimental measurements and the modeling framework. A user-contributed database contains a set of tools for easy input, search and modification of thousands of records. The OCHEM database is based on the wiki principle and focuses primarily on the quality and verifiability of the data. The database is tightly integrated with the modeling framework, which supports all the steps required to create a predictive model: data search, calculation and selection of a vast variety of molecular descriptors, application of machine learning methods, validation, analysis of the model and assessment of the applicability domain. Our intention is to make OCHEM a widely used platform to perform the QSPR/QSAR studies online and share it with other users on the Web.

By such integration, scientists can model In vitro assays using In silico descriptor packages while making benefit of multi-learning features and automatics applicability domain estimation.

http://ochem.eu

Cite abstracts in this supplement using the relevant abstract number, e.g.: Abdelaziz *et al.*: QSAR modeling for In vitro assays: linking ToxCast™ database to the integrated modeling framework "OCHEM". *Journal of Cheminformatics* 2012, **4**(Suppl 1):P62